NUSC Technical Report 7064
15 October 1991

Feature Detection for Model Assessment in State Estimation

D. J. Ferkinhoff
J. G. Baylog
K. F. Gong
Combat Control Systems Department

S. C. Nardone
University of Massachusetts Dartmouth

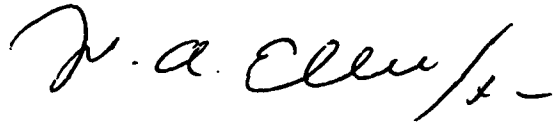
Naval Underwater Systems Center
Newport, Rhode Island • New London, Connecticut

PREFACE

This work was conducted under the Submarine Contact Management Project of the Combat Control Block Program, with support under Contract N66604-90-D104-0003. The NUSC principal investigator is K.F. Gong (Code 2211). The sponsoring activity is the Chief of Naval Research, Office of Naval Technology, program manager D. C. Houser (OCNR-232).

The technical reviewer for this report was V. J. Aidala (Code 2211).

REVIEWED AND APPROVED: 15 OCTOBER 1991



P. A. LaBrecque
Head, Combat Control Systems Department

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	15 October 1991	

4. TITLE AND SUBTITLE

Feature Detection for Model Assessment in State Estimation

5. FUNDING NUMBERS

6. AUTHOR(S)

D. J. Ferkinhoff S. C. Nardone*
J. G. Baylog K. F. Gong

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Underwater Systems Center
Newport Laboratory
Newport, Rhode Island 02841-5047

8. PERFORMING ORGANIZATION REPORT NUMBER

TR 7064

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Chief of Naval Research
Office of Naval Technology (OCNR-232)
Arlington, VA 22203

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

*S. C. Nardone is affiliated with the University of Massachusetts Dartmouth, North Dartmouth, MA 02747.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release;
distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The presence of deterministic features in a residual sequence is oftentimes an indication of a modeling error in the state estimation process. Here, three methods of detecting and extracting the features of "jump" and/or "drift" in a predicted residual sequence are developed. Two of the methods are traditional ones. The first is a multiple hypothesis, generalized, likelihood ratio test that results in a chi-squared (XSQ) test statistic. The second is a similar, but computationally more efficient, intuitively derived test resembling a modified Neyman-Pearson (MNP) test. The third method is a nontraditional one; it uses a backpropagation artificial neural network trained to emulate the MNP test. Monte Carlo experimental results show that the XSQ and MNP give essentially identical results, while the ANN -- although apparently outperforming the XSQ in feature detection -- does so at the expense of a higher feature misclassification probability, which is an undesirable effect. Overall, the ANN is judged as a feasible approach to feature detection, and improved performance is expected with better network training.

14. SUBJECT TERMS

Underwater Tracking
Target Motion Analysis
Statistical Theory
Artificial Neural Networks
Mathematical Models

15. NUMBER OF PAGES
53

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

TABLE OF CONTENTS

Section		Page
	LIST OF ILLUSTRATIONS	ii
	LIST OF TABLES	ii
1.	INTRODUCTION	1
2.	PROCESS MODEL	3
3.	BINARY HYPOTHESIS TEST	11
3.1	Likelihood Ratio Test (LRT)	11
3.2	Generalized Likelihood Ratio Test (GLRT)	19
3.3	Artificial Neural Network (ANN) Hypothesis Test	25
3.4	Training the Artificial Neural Network	27
4.	MULTIPLE HYPOTHESIS TEST	29
4.1	Likelihood Ratio Test	29
4.2	Modified Neyman-Pearson Test	37
4.3	Artificial Neural Network	38
5.	EXPERIMENTAL RESULTS AND DISCUSSION	39
6.	SUMMARY AND CONCLUSIONS	47
	REFERENCES	48

LIST OF ILLUSTRATIONS

Figure		Page
1	State Estimation Process	1
2	Binary LRT	16
3	LRT Receiver Operating Characteristic	16
4	Binary GLRT	23
5	GLRT Receiver Operating Characteristic	24
6	Typical Single Neuron	25
7	Typical Feedforward Network with Two Hidden Layers	26
8	Illustration of the Sigmoid Function	28
9	Illustration of Network Training	28
10	ANN Receiver Operating Characteristic (Jump Only)	41
11	Theoretical GLRT Receiver Operating Characteristic (Jump Only)	41
12a	χSQ Receiver Operating Characteristic (Jump)	42
12b	ANN Receiver Operating Characteristic (Jump)	42
13a	χSQ Jump Output Given Drift Only	43
13b	χSQ Jump Output Given Jump Only	43
13c	χSQ Jump Output Given Drift with Same Polarity	44
13d	χSQ Jump Output Given Drift with Opposite Polarity	44
14a	ANN Jump Output Given Drift Only	45
14b	ANN Jump Output Given Jump Only	45
14c	ANN Jump Output Given Drift with Same Polarity	46
14d	ANN Jump Output Given Drift with Opposite Polarity	46

LIST OF TABLES

Table		Page
1	Multiple Hypothesis Test	37

FEATURE DETECTION FOR MODEL ASSESSMENT IN STATE ESTIMATION

1. INTRODUCTION

This report addresses the problem of feature detection for the purpose of assessing the validity of models used in state estimation. The state estimation process is depicted schematically in figure 1. Here, the residuals — the difference between the observed measurements and estimates of the measurements (based on the state estimate) — are mapped into a correction term for the state estimate through a variant of the gradient equations relating the measurements to the state. The estimation algorithm converges to an estimate that provides a "best fit" to the observed measurements. When the system model sufficiently reflects the actual process by which the measurements were produced, the residuals are noise-like in character. That is, they are devoid of any deterministic features.

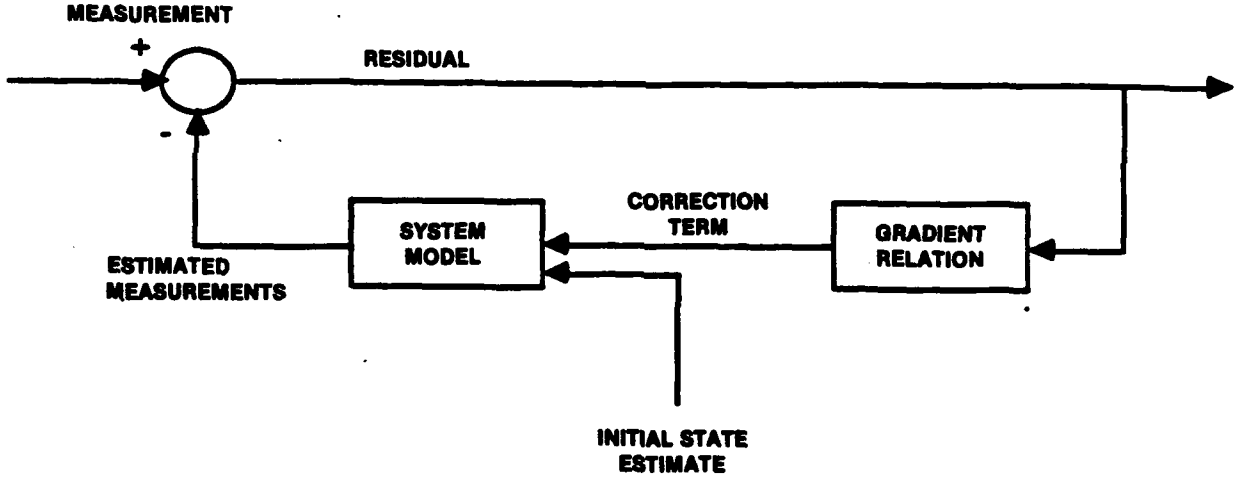


Figure 1. State Estimation Process

Oftentimes, indications of errors in the system model manifest themselves as deterministic features in the residuals. Consequently, the processing of measurement data for state estimation in an uncertain modeling environment requires continuous scrutiny of measurements and measurement residuals. Once a feature, or set of features, has been detected, appropriate techniques can be implemented to compensate or identify and correct for the modeling error. One such system, TARSIA, has been considered previously for model assessment (reference 1).

In TARSIA, the detected residual features are interpreted in an evidential reasoning system to identify the possible modeling anomaly. The work presented in this report extends the feature detection capability of TARSIA.

Specifically, this report details three methods to detect and extract the features of jump and/or drift in a predicted residual sequence. Two of the methods, a generalized likelihood chi-squared(χSQ) test and a modified Neyman-Pearson (MNP) test, are classical numerical approaches. The third method uses an artificial neural network (ANN) and is a nontraditional approach. The likelihood ratio test (LRT) and generalized likelihood ratio test (GLRT) for the binary hypothesis test are developed to provide a background for future derivations. Additionally, a backpropagation ANN is trained to emulate the LRT for the binary hypothesis test. Building on the binary test leads to development of the more complicated multiple hypothesis GLRT, which is shown to result in a χSQ statistic. An intuitive approach, based on the χSQ test, results in an MNP test that is more computationally efficient than the χSQ test. Because of this computational efficiency, the multiple hypothesis ANN is trained to emulate the MNP test.

Experimental results, obtained from Monte Carlo simulations using synthetic data, are presented in the form of receiver operating characteristic (ROC) curves. For the binary hypothesis test, the ANN and GLRT give nearly identical results and are in good agreement with the theoretically predicted performance. For the multiple hypothesis test, the χSQ test and the MNP test give essentially identical results, while the ANN performance departs somewhat from that of the conventional tests. Although the ANN outperforms the χSQ test in the probability of feature detection, it does so at the expense of a higher feature misclassification (i.e., detecting feature A when feature B is present), which is undesirable. Overall, the performance of the ANN demonstrates the feasibility of the technique but also indicates that the training of the network may have been incomplete.

The remainder of this report is organized into five sections. Section 2 defines the process and mathematical models under consideration. Section 3 presents the binary hypothesis test as background and discusses the ANN implemented for that problem. Section 4 extends the binary hypothesis test to multiple hypotheses, develops the χSQ test, presents the MNP test, and discusses the multiple-feature ANN. Section 5 presents the experimental results and discusses their implications. Section 6 offers a summary and conclusions.

2. PROCESS MODEL

Consider the following discrete time dynamic state model that generates measurements z(k):

$$x(k+1) = \Phi x(k) + Bu(k) + \Gamma w(k), \tag{1a}$$

$$z(k) = H[x(k)] + n(k). \tag{1b}$$

Here, x(k) is the Kx1 dimensional state vector; u is the input vector; w is zero-mean, white, Gaussian process noise; z is the measurement; and n is zero-mean, white, Gaussian measurement noise that is independent of the process noise. The matrices Φ, B, Γ are, respectively, the state transition matrix, the input matrix, and the process noise input matrix. The function H is the nonlinear function relating the state to the measurement, and k is the time index. The state x(k) is available only as an estimate $\hat{x}(k|k)$, along with the estimated error covariance matrix P(k|k). The estimate $\hat{x}(k|k)$ of the state at time k is based on all the past data up to and including time k. The estimate is available from a suitable estimator like an extended Kalman filter or maximum likelihood estimator.

The residual $\tilde{z}(k|k-1)$ is the difference between the measurement z(k) and the estimate of the measurement $\hat{z}(k|k-1)$; that is,

$$\tilde{z}(k|k-1) = z(k) - \hat{z}(k|k-1), \tag{2a}$$

where

$$\hat{z}(k|k-1) = H[\hat{x}(k|k-1)], \tag{2b}$$

$$\hat{x}(k|k-1) = \Phi \hat{x}(k-1|k-1) + Bu(k-1), \tag{2c}$$

where $\hat{x}(k|k-1)$ is the state estimate at time k based on the measurements up to and including time k - 1 but not time k. The process to be investigated here is the predicted residual sequence

$$\tilde{Z} = [\tilde{z}_1, \tilde{z}_2, ..., \tilde{z}_N]^T, \tag{3a}$$

where

$$\tilde{z}_i = \tilde{z}(k+i|k), \tag{3b}$$

$$\tilde{z}(k+i|k) = H[\hat{x}(k+i|k)], \tag{3c}$$

$$\hat{x}(k+i|k) = \Phi^i \hat{x}(k|k) + \sum_{j=1}^{i} \Phi^{(i-j)} Bu(k+j-1), \tag{3d}$$

where $\hat{x}(k+i|k)$ is produced by propagating the state estimate at time k forward in time without updating the estimate.

Examining the i^{th} element of the predicted residual sequence

$$\tilde{z}(k+i|k) = H[x(k+i)] - H[\hat{x}(k+i|k)] + n(k+i), \tag{4}$$

and expanding the measurement function in a Taylor series about the estimate, with

$$\delta x(k+i) = \hat{x}(k+i|k) - x(k+i), \tag{5}$$

yields the approximate expression

$$\tilde{z}(k+i|k) \approx - a^T[\hat{x}(k+i|k)]\delta x(k+i) + n(k+i), \tag{6a}$$

where

$$a[\hat{x}(k+i|k)] = \left. \frac{\delta H[x(k+i)]}{\delta x(k+i)} \right|_{x(k+i) = \hat{x}(k+i|k)} \tag{6b}$$

The dynamic equation for the propagation of the state estimation error in terms of the fixed error at time k is

$$\delta x(k+i) = \Phi^i \delta x(k) - g^T(k+i)W, \tag{7a}$$

where

$$g^T(k+i) = [\Phi^{i-1}\Gamma, \Phi^{i-2}\Gamma, ..., \Phi\Gamma, \Gamma, 0, ..., 0], \tag{7b}$$

and

$$W^T = [w(k), w(k+1), ..., w(k+i-2), w(k+i-1), w(k+i), ..., w(K-1)]. \tag{7c}$$

The entire predicted residual sequence can be approximated as

$$\tilde{Z} = -A[\hat{x}(k|k)]\delta x(k) - GW + N_K, \tag{8a}$$

where

$$A\,[\hat{x}(k|k)] = \left[\begin{array}{c} \vdots \\ \left(\dfrac{\partial H[x(k+i)]}{\partial x(k)}\right)^{T} \\ \vdots \end{array}\right]\Bigg|_{x(k)=\hat{x}(k|k)} = \left[\begin{array}{c} \vdots \\ a^{T}[\hat{x}(k+i|k)]\Phi^{i} \\ \vdots \end{array}\right], \tag{8b}$$

and

$$N_K = \left[\begin{array}{c} \vdots \\ n(k+i) \\ \vdots \end{array}\right], \quad G = \left[\begin{array}{c} \vdots \\ g^{T}(k+i) \\ \vdots \end{array}\right]. \tag{8c}$$

The predicted residual sequence can be approximated as a jointly Gaussian random vector. For unbiased state estimates, $\tilde{Z}$ is zero mean and its probability density function is

$$P(\tilde{Z}) = \left[(2\pi)^{K/2} \det |V|^{1/2}\right]^{-1} \exp\left[-\frac{1}{2}\tilde{Z}^{T}V^{-1}\tilde{Z}\right]. \tag{9a}$$

Here, V is the covariance matrix of the predicted residual vector defined as

$$\mathrm{cov}\,(\tilde{Z}) = E\,[\tilde{Z}\tilde{Z}^{T}] - E[\tilde{Z}]\,E[\tilde{Z}^{T}], \tag{9b}$$

$$\mathrm{cov}\,(\tilde{Z}) = \mathcal{E}[(N_K - A\delta x - GW)(N_K - A\delta x - GW)^{T}]. \tag{9c}$$

Thus,

$$V = AP(k|k)A^{T} + S^{2} + GQG^{T}, \tag{9d}$$

where

$$S^{2} = E[N_K N_K^{T}], \tag{9e}$$

$$P(k|k) = E[\delta x(k)\delta x^T(k)], \tag{9f}$$

$$Q = E[WW^T]. \tag{9g}$$

Note that the presence of process noise in the predicted residual sequence results in an effective noise

$$N_e = N_K - GW, \tag{10a}$$

with an increased covariance

$$S_e = S^2 + GQG^T. \tag{10b}$$

Further, it is noted that the probability density function for the predicted residuals is only approximately Gaussian. However, the approximation remains valid whenever the second-order and higher terms of the Taylor series expansion of the measurement nonlinearity can be ignored. Thus, two assumptions on the predicted residual sequence are invoked to ensure the reasonableness of the Taylor series approximation. First, the dynamic model of equation (1), which is used to obtain the estimate, must be a reasonably accurate description of the real process. Second, the state error, and its effect over the time interval of interest, must remain sufficiently small to allow the validity of a first-order Taylor series approximation of the predicted residual sequence. These two assumptions are the basis of the null hypothesis, i.e.,

H_o: No modeling error and a "good" state estimate.

Under the H_o hypothesis, the predicted residual process is a zero-mean, jointly Gaussian random vector; that is, the probability density function is

$$P(\tilde{Z}|H_o) = \left[(2\pi)^{K/2} \det|V|^{1/2}\right]^{-1} \exp\left[-\frac{1}{2}\tilde{Z}^T V^{-1}\tilde{Z}\right], \tag{11}$$

where V is the covariance matrix given by equation (9d). Note that although elements of the predicted residual vector are correlated, there always exists a set of basis vectors in which $\tilde{Z}$ can be represented by uncorrelated (and hence independent) components. For many applications considered here, both process noise and state estimation error covariances are small compared with the measurement noise covariance. For these cases, the elements of $\tilde{Z}$ are also independent.

Under the alternate hypothesis H_1, there is a modeling error. The original state and measurement models of equations (1) do not adequately describe the system dynamics. Modeling errors may arise from unknown system inputs, a change in the measurement model, or an unobserved change in the measurement noise. Examples of such modeling anomalies are, respectively, a target maneuver, change in propagation path, and measurement bias. Regardless of

the cause of the modeling error, its observable effect is often a feature in the predicted residual sequence.

Under the mismodeling hypothesis H_1, the predicted residuals exhibit features that are distinguishable from those of hypothesis H_0. For hypothesis H_1, a $(q-1)^{th}$ order polynomial is used to model the anomalous features of the predicted residuals, that is,

$$H_1: \quad \tilde{Z}(k+i|k) = -a^T[\hat{x}(k+i|k)]\delta\hat{x}(k) + m(k+i) + n(k+i) - g^T(k+i)W\,, \tag{12a}$$

where

$$\begin{aligned} m(k+i) = [&a_0 + a_1(t_{k+i} - t_0) + a_2(t_{k+i} - t_0)^2 + \ldots \\ &+ a_{q-1}(t_{k+i} - t_0)^{q-1}]u(t_{k+i} - t_0), \end{aligned} \tag{12b}$$

$$u(t) = \begin{cases} 0 & t < 1 \\ 1 & t > 1 \end{cases},$$

and t_0 is the time of the modeling anomaly. The predicted residual sequence under H_1 is

$$H_1: \quad \tilde{Z} = -A[\hat{x}(k|k)]\delta\hat{x}(k) - GW + M_1 + N_K, \tag{13a}$$

where

$$M_1 = B\,\bar{m}, \tag{13b}$$

$$B = \begin{bmatrix} 0 & \\ \vdots & \\ 1 & 0 \\ 1 & (t_1 - t_0)\ (t_1 - t_0)^2 \ldots \\ & \vdots \end{bmatrix},$$

and

$$\bar{m} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{q-1} \end{bmatrix}.$$

The term M_1 contributes to the mean value of $\tilde{Z}$; hence,

$$E[\tilde{Z}] = M_1, \tag{14a}$$

and

$$\text{cov}\, E[\tilde{Z}] = E[(\tilde{Z} - M_1)(\tilde{Z} - M_1)^T] = V. \tag{14b}$$

Thus, the covariance under both assumptions is the same and only the means differ. The probability density function (PDF) for the jointly Gaussian random vector under H_1 is

$$P(\tilde{Z}|H_1) = \left[(2\pi)^{K/2} \det |V|^{1/2}\right]^{-1} \exp\left[-\frac{1}{2}(\tilde{Z} - M_1)^T V^{-1}(\tilde{Z} - M_1)\right]. \tag{15}$$

Similar results can be developed for both the smoothed and filtered residuals. For the smoothed residual process, the state estimation error is correlated with the past measurement noise. The conditional PDF or likelihood function for the observations (no process noise)

$$Z = [z_1, ..., z_K]^T, \tag{16}$$

given the state x, is

$$P(Z|x) \Rightarrow N[Z(x), S^2], \tag{17a}$$

or

$$P(Z|x) = \left[(2\pi)^{K/2} \det |S^2|^{1/2}\right]^{-1} \exp\left[-\frac{1}{2}[Z - Z(x)]^T S^{-2}[Z - Z(x)]\right], \tag{17b}$$

where

$$Z(x) = \left[H[x(1)], ..., H[x(K)]\right]. \tag{17c}$$

The maximum likelihood estimate is the value of the state x that maximizes the likelihood function of equation (17). Taking the logarithm of the likelihood function yields the log-likelihood equation:

$$\ln[P(Z|x)] = [\text{constant}] + \frac{1}{2}[Z - Z(x)]^T S^{-2} [Z - Z(x)] . \tag{18}$$

Differentiating the log-likelihood equation with respect to the state at time k and setting the result to zero, one can obtain the maximum likelihood estimate by solving the equation

$$A^T[x(k)]S^{-2}\left[Z - Z[x(k)]\right]\Big|_{x(k) = \hat{x}(k|k)} = 0 . \tag{19}$$

For the linearized smoothed residuals,

$$\tilde{Z}_S = -A[\hat{x}(k|k)]\delta x(k) + N_K, \tag{20}$$

where A[ ] is defined in equation (8b). The error in the estimate $\delta x(k)$ is approximately

$$\delta x(k) \approx [A^T S^{-2} A]^{-1} A^T S^{-2} N_K, \tag{21a}$$

and the state error covariance is

$$P(k|k) = E\{[\delta x(k)\delta x^T(k)]\}, \tag{21b}$$

which can be approximated as

$$P(k|k) \approx [A^T S^{-2} A]^{-1}. \tag{21c}$$

Consequently, from equation (9), the PDF for the smoothed residuals under H_0 is

$$P(\tilde{Z}|H_0) \Rightarrow N(0, V_S), \tag{22a}$$

where

$$V_S = \text{cov}(\tilde{Z}_S),$$

or

$$V_S = E[(N_K - A\delta x)(N_K - A\delta x)^T];$$

substituting from equation (21a) for $\delta x(k)$ yields

$$V_S = S^2 - A^T P(k|k)A. \tag{22b}$$

Under hypothesis H_1 (i.e., a modeling anomaly exists), the PDF of the predicted residuals is (possibly) nonzero-mean jointly Gaussian:

$$P(\tilde{Z}|H_1) \Rightarrow N(M_1,V), \tag{23}$$

where M_1 is the mean vector defined in equation (13), V is the covariance matrix given by equation (9d), and $N(M_1,V)$ is the Gaussian PDF with mean M_1 and covariance V. Similarly, under hypothesis H_1 for the smoothed residuals,

$$P(\tilde{Z}_S|H_1) \Rightarrow N(M_1,V_S). \tag{24}$$

When the filtered residual sequence is obtained from the extended Kalman filter, the estimate $\hat{x}(k|k)$ is the conditional mean, and the innovations process is white (reference 2). For this process, the PDFs are

$$H_0: \quad P(\tilde{Z}_f|H_0) \Rightarrow N(0,V_f), \tag{25a}$$

$$H_1: \quad P(\tilde{Z}_f|H_1) \Rightarrow N(M_1,V_f), \tag{25b}$$

where

$$V_f = \text{diag}\,[a^T P(i|i)a + \sigma_n^2]. \tag{25c}$$

The above development characterizes three processes: predicted residuals, smoothed residuals, and filtered (conditional mean) residuals. The predicted residuals are obtained from future measurements and the state prediction; the smoothed residuals are obtained by fitting the past data with the best state estimate; the filtered residuals, usually obtained from a Kalman filter, involve the current state estimate and the next measurement. While the filtered innovations obtained from a Kalman filter are a white process, both the smoothed and predicted residuals are correlated to some degree; however, for small state estimation error, they can often be approximated as a white process. In this investigation, attention is focused on the predicted residuals, and for some results is limited to the case of negligible state estimation error.

3. BINARY HYPOTHESIS TEST

The optimum hypothesis test, for either Bayes or Neyman-Pearson criteria, is the likelihood ratio test. In this section the two binary hypothesis tests that decide if a modeling anomaly is present are presented. They are the likelihood ratio test (LRT) and generalized likelihood ratio test (GLRT). The LRT is applicable to detection of known signals, while the GLRT is appropriate for the detection of a known signal form having unknown values. Simple examples of the detection of signals in noise are given along with experimental results. The design, implementation, and testing of an artificial neural network (ANN) for the binary hypothesis test are also presented. The development of the binary hypothesis test of this section is a precursor to the multiple hypothesis test of the next section.

3.1 LIKELIHOOD RATIO TEST (LRT)

The (binary) LRT compares the ratio of likelihood functions from two hypotheses to a threshold and selects one of the two hypotheses as being true. The predicted residual probability density functions (PDFs) from section 2 are used as the likelihood functions. The two hypotheses are

H_0: no modeling anomaly and a "good" state estimate ($M_1 = 0$),

H_1: presence of a modeling anomaly ($M_1 \neq 0$).

Under these two hypotheses, the likelihood ratio

$$\frac{P(\tilde{Z}|H_1)}{P(\tilde{Z}|H_0)} \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda \tag{26}$$

is formed. $P(\tilde{Z}|H_0)$ and $P(\tilde{Z}|H_1)$ were defined in equations (11) and (15), respectively. The constant λ is the threshold. Four outcomes from this test are possible: two correct decisions of selecting H_0 when H_0 occurred or H_1 when H_1 occurred, and two incorrect decisions of selecting H_1 when H_0 occurred (false detection or false alarm) or H_0 when H_1 occurred (missed-detection). Note that since $P(\tilde{Z}|H_0)$ and $P(\tilde{Z}|H_1)$ are both positive, it follows that $\lambda > 0$. Further, if H_1 is more likely than H_0, the ratio on the left-hand side of the inequality is greater than one; and, conversely, if H_0 is more likely than H_1, the ratio is less than one.

Typically, λ is set to achieve a specified probability of false alarm, from which a detection probability follows. Thus, the development here focuses on determining false alarm and detection probabilities.

Substituting for $P(\tilde{Z}|H_0)$ and $P(\tilde{Z}|H_1)$ from the previous section and reducing the ratio yields

$$\exp\left[-\frac{1}{2}\|\tilde{Z} - M_1\|^2_{V^{-1}} + \frac{1}{2}\|\tilde{Z}\|^2_{V^{-1}}\right] \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda. \tag{27}$$

Here, $\tilde{Z}$ is the predicted residual sequence, and M_1, the modeling anomaly, is the signal to be detected. Expanding the weighted-norms of the exponent and reducing yields

$$\exp\left[-\frac{1}{2}\{M_1^T V^{-1} M_1 - 2M_1^T V^{-1}\tilde{Z}\}\right] \underset{H_1}{\overset{H_0}{\lessgtr}} 1. \tag{28}$$

Because the logarithm is a monotonic function, the logarithm of the likelihood ratio can be used without altering the outcome. Taking the natural logarithm of equation (28) yields

$$M_1^T V^{-1}\tilde{Z} - \frac{1}{2}M_1^T V^{-1} M_1 \underset{H_1}{\overset{H_0}{\lessgtr}} \ln\lambda. \tag{29}$$

If the signal M_1 is known, then the second term on the left of inequality (29) can be moved to the right-hand side, resulting in the LRT:

$$M_1^T V^{-1}\tilde{Z} \underset{H_1}{\overset{H_0}{\lessgtr}} \ln\lambda + \frac{1}{2}M_1^T V^{-1} M_1. \tag{30}$$

The left-hand side of inequality (30) is the LRT test statistic. Because it is a linear combination of Gaussian random variables, it is also a Gaussian random variable. Further, because the LRT depends solely on the test statistic,

$$\ell(\tilde{Z}) = M_1^T V^{-1} \tilde{Z}; \tag{31}$$

then, $\ell(\tilde{Z})$ is a sufficient statistic for LRT. Thus, knowledge of the probability density function for $\ell(\tilde{Z})$ is sufficient for determining the probability of false alarm and the probability of detection. Under the hypothesis H_0: $M_1 = 0$,

$$E\ [\ell(\tilde{Z}|H_0)] = E\left[M_1^T V^{-1}[A\delta x(k) + N_K]\right] = 0. \tag{32}$$

The variance of $\ell(\tilde{Z}|H_0)$ is

$$\text{var}\ [\ell(\tilde{Z}|H_0)] = E\left[\{M_1^T V^{-1}[A\delta x(k) + N_K]\}\ \{M_1^T V^{-1}[A\delta x(k) + N_K]\}^T\right], \tag{33a}$$

which reduces to

$$\text{var}\ [\ell(\tilde{Z}|H_0)] = M_1^T V^{-1} M_1. \tag{33b}$$

Define the variance of $\ell[(\tilde{Z}|H_0)]$ as

$$\sigma_\ell^2 = M_1^T V^{-1} M_1. \tag{34}$$

Similarly, for the alternate hypothesis H_1: $M_1 \neq 0$,

$$E[\ell(\tilde{Z}|H_1)] = E\left[M_1^T V^{-1}[A\delta x(k) + N_K + M_1]\right],$$

which reduces to

$$E[\ell(\tilde{Z}|H_1)] = M_1^T V^{-1} M_1,$$

or

$$E[\ell(\tilde{Z}|H_1)] = \sigma_\ell^2. \tag{35}$$

The variance of $\ell(\tilde{Z}|H_1)$ is also σ_ℓ^2. The PDFs for $\ell(\tilde{Z})$ under both hypotheses are

$$P(\ell|H_0) \Rightarrow N(0, \sigma_\ell^2), \tag{36a}$$

$$P(\ell|H_1) \Rightarrow N(\sigma_\ell^2, \sigma_\ell^2). \tag{36b}$$

The LRT of equation (30) can be rewritten as

$$\ell(\tilde{Z}) \underset{H_1}{\overset{H_0}{\lessgtr}} \varepsilon_{LRT}, \tag{37a}$$

where the threshold

$$\varepsilon_{LRT} = \ln\lambda + \frac{1}{2} M_1^T V^{-1} M_1 . \tag{37b}$$

The probability of a false alarm P_F is the probability that the test statistic $\ell(\tilde{Z})$ exceeds the specified threshold ε_{LRT} when the correct hypothesis is H_0. Thus, P_F is the area under $P(\ell|H_0)$ for $\ell > \varepsilon$, or

$$P_F = \int_{\varepsilon_{LRT}}^{\infty} P(\ell|H_0) d\ell; \tag{38}$$

or, since $P(\ell|H_0)d\ell \Rightarrow N(0, \sigma_\ell^2)$,

$$P_F = 1 - \text{erf}\,[\varepsilon_{LRT}/\sigma_\ell] , \tag{39}$$

where erf (*) is the standard error function,

$$\text{erf}[x] = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x} e^{-x^2} dx . \tag{40}$$

In an analogous manner, the probability of detecting the signal P_D is the probability that the test statistic is greater than the threshold, given that the correct hypothesis is H_1. Thus,

$$P_D = \int_{\varepsilon_{LRT}}^{\infty} P(\ell|H_1)d\ell, \tag{41}$$

which for the Gaussian case is

$$P_D = 1 - \text{erf}\,[(\varepsilon_{LRT} - \sigma_\ell^2)/\sigma_\ell]. \tag{42}$$

Figure 2 illustrates the P_F and P_D of equations (39) and (42).

Given a desired P_F, the corresponding threshold can be obtained from a table or computer program. The P_D for a given ε is determined by the system parameter σ_ℓ^2, which is a function of signal design (M_1) and noise V. Because σ_ℓ^2 is a measure of signal-to-noise ratio (SNR), the performance of a hypothesis test as a signal detector is often characterized as P_D vs P_F for a given SNR. These curves are shown in figure 3 and are called *receiver operating characteristic (ROC)* curves (reference 3). Although the specific case of the Gaussian test statistic is treated here, as given in equations (39) and (42) and illustrated in figures 2 and 3, the results of (38) and (41) hold for arbitrary densities of $\ell(\tilde{Z})$.

A case of special interest is examined below. Presently, the aspect of state estimation error is neglected resulting in a simple signal detection problem. In all cases the predicted residual covariance is $V = \sigma^2 I$. For the case of interest, under the two hypotheses

$$H_0: \tilde{Z} = N_K,$$

$$H_1: \tilde{Z} = M_{i+1} + N_K, \quad i \geq 0,$$

where $\tilde{Z}$ is the predicted residual, which with $\delta x = 0$ implies $V = \sigma^2 I$, and the signal M_{i+1} is

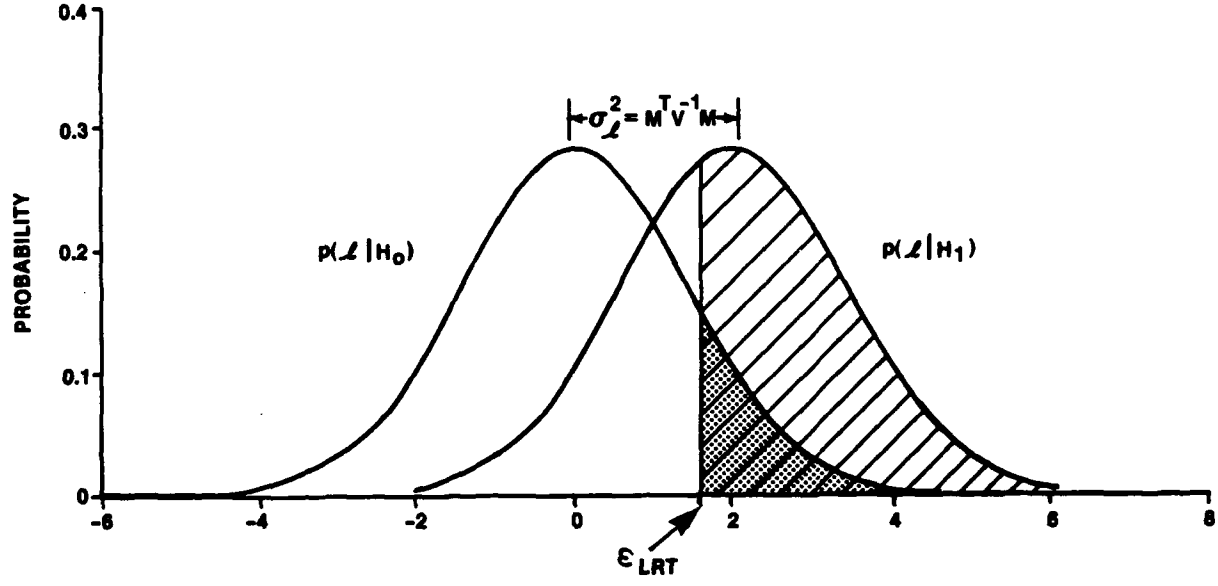


Figure 2. Binary LRT

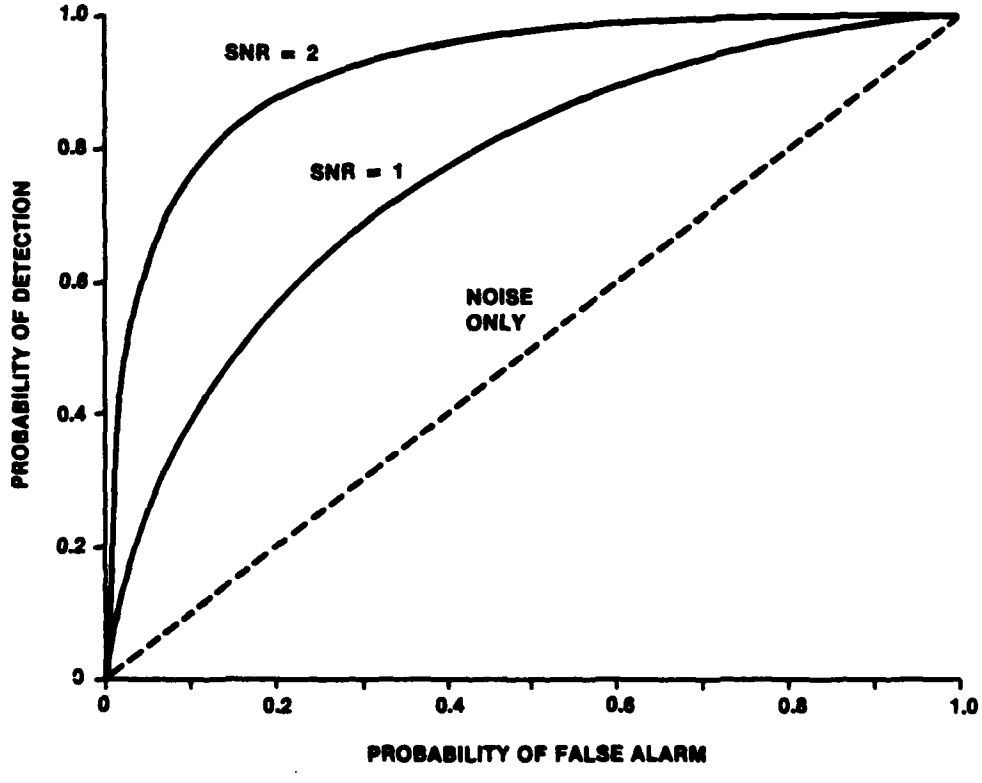


Figure 3. LRT Receiver Operating Characteristic

$$M_{i+1} = a_i \begin{bmatrix} \vdots \\ [(j-1)T]^i \\ \vdots \end{bmatrix}, \tag{43}$$

where j is the j^{th} time index, T is the sampling period, and "a" is a known coefficient. Substituting equation (43) into the LRT of equation (30) yields

$$\left(\frac{a_i T^i}{\sigma_n^2} \sum_{j=1}^{K} (j-1)^i \tilde{z}_j \right) \underset{H_1}{\overset{H_0}{<>}} \ln\lambda + \frac{1}{2} \frac{a_i^2 T^{2i}}{\sigma_n^2} \sum_{j=1}^{K} (j-1)^{2i}. \tag{44}$$

Normalizing equation (44) by the term

$$k_{ni} = \frac{a_i T^i}{\sigma_n / \sqrt{k_{2i}}}, \tag{45a}$$

where

$$k_{2i} = \sum_{j=1}^{K} (j-1)^{2i}, \tag{45b}$$

results in a test statistic for the i^{th} feature defined as

$$f_i(\tilde{Z}) = \left((\sigma_n \sqrt{k_{2i}})^{-1} \sum_{j=1}^{K} (j-1)^i \tilde{z}_j \right) \underset{H_1}{\overset{H_0}{<>}} \frac{\ln\lambda}{k_{ni}} + \frac{1}{2} k_{ni}. \tag{46}$$

The new test statistic $f_i(\tilde{Z})$ is a univariant Gaussian random variable. Consequently, the test statistic PDFs under the two hypotheses are

$$P(\ell_i \mid H_0) \Rightarrow N(0,1), \tag{47a}$$

$$P(\ell_i \mid H_1) \Rightarrow N(k_{ni},1). \tag{47b}$$

Note that the normalizing factor of equation (45a) is the square-root of the last term on the right-hand side of equation (44) and is the ratio of signal amplitude to effective noise standard deviation of the smoothing interval. The cases of interest are for i = {0,1,2,}. In these cases, the test statistics are

$$\ell_0(\tilde{Z}) = \left[\sigma_n\sqrt{K}\right]^{-1} \sum_{j=1}^{K} \tilde{z}_j, \tag{48a}$$

$$\ell_1(\tilde{Z}) = \left[\sigma_n\sqrt{K(K-1)(2K-1)/6}\right]^{-1} \sum_{j=1}^{K} (j-1)\,\tilde{z}_j, \tag{48b}$$

$$\ell_2(\tilde{Z}) = \left[\sigma_n\sqrt{K(2K-1)(3K^2-3K-1)(K-1)/30}\right]^{-1} \sum_{j=1}^{K} (j-1)^2\,\tilde{z}_j, \tag{48c}$$

and the normalizing factors (mean of (47b) as defined in (45a)) are

$$k_{no} = \frac{a_o}{\sigma_n/\sqrt{K}}, \tag{49a}$$

$$k_{n1} = \frac{a_1 T}{\sigma_n/\sqrt{K(K-1)(2K-1)/6}}, \tag{49b}$$

$$k_{n2} = \frac{a_2 T^2}{\sigma_n/\sqrt{K(K-1)(2K-1)(3K^2-3K-1)/30}}. \tag{49c}$$

These three cases are examples of a signal that exhibits the features of a step (jump), ramp (drift), and simple quadratic (curvature).

3.2 GENERALIZED LIKELIHOOD RATIO TEST (GLRT)

In the previous case of the LRT, the signal to be detected was completely known. However, for some signal detection problems, a parameter of the signal (such as signal amplitude) is unknown and may vary over a set of values. This type of detection problem is referred to as a composite hypothesis. A hypothetical test can be constructed for the composite hypothesis using the correct (but unknown) value of the signal parameter in the design of the optimal LRT. This test is an upperbound on the performance of any other test. A uniformly most powerful (UMP) test has a P_D that is greater than or equal to any other test for a given P_F. Thus, if a test's actual performance achieves the bound obtained by the LRT using the correct signal parameter, then it is a UMP test. For a UMP test to exist, it is both necessary and sufficient that a likelihood ratio test for every value of the signal parameter can be constructed without knowledge of the signal parameter (reference 3).

Consider the polynomial case of interest from the preceding section. The LRT is given by equation (44), which is rewritten here in a slightly different form:

$$\ell_1(\tilde{Z}) = \sum_{j=1}^{K} (j-1)^i \; \tilde{z}_j \underset{H_1}{\overset{H_0}{\lessgtr}} \frac{\varepsilon\sigma_n^2}{a_i T^i}, \qquad a_i > 0. \tag{50}$$

Here, the scaling terms on $\ell(\tilde{Z})$ that rendered it univariant have been moved to the other side of the equation. In equation (50), the unknown signal parameter a_i has been assumed to be positive. Clearly, if a_i is negative then the inequality signs must be reversed. Thus, although the magnitude of a_i is not required to construct the LRT, its sign must be known to properly construct the test. Therefore, the LRT cannot be implemented without knowledge of the polarity of the signal parameter, and since the polarity of the feature parameters that may be present in a residual sequence cannot be predicted, a UMP test does not exist for this problem. Consequently, an alternative to the UMP test must be used.

An alternate approach is to estimate the value of the signal parameter and use the estimated parameter in the LRT. This is the generalized likelihood ratio test (GLRT). The signal model is defined by equations (8) and (9) for H_0 and by equations (13) and (15) for H_1. The unknown signal parameters $\bar{m}$ are defined in equation (13b). Note that hypothesis H_0 is just the special case that $\bar{m}$ has the specific value of zero. The maximum likelihood estimate of $\bar{m}$ is

$$\hat{\bar{m}} = \left[B^T V^{-1} B\right]^{-1} B^T V^{-1} \tilde{Z}, \tag{51}$$

and the estimate of the predicted residual mean M is

$$\hat{M} = B\,\hat{\bar{m}}. \tag{52}$$

Substituting $\hat{M}$ for M_1 of equation (29) (the subscript is being suppressed) yields

$$\hat{M}^T V^{-1} \tilde{Z} - \frac{1}{2}\hat{M}^T V^{-1}\hat{M} \underset{H_1}{\overset{H_0}{\lessgtr}} \ln\lambda. \tag{53}$$

Substituting equation (51) into equation (52) and using the result in equation (53) yields, after some manipulation,

$$\ell(\tilde{Z}) = \tilde{Z}^T W \tilde{Z} \underset{H_1}{\overset{H_0}{\lessgtr}} \ln\lambda^2, \tag{54}$$

where $W = V^{-1}B[B^TV^{-1}B]^{-1}B^TV^{-1}$. Note that the test statistic is a weighted norm of $\tilde{Z}$ and is positive; thus, λ must be greater than 1 for a meaningful test.

Inserting $I = [B^TV^{-1}B]^{-1}[B^TV^{-1}B]$ into the weight, factoring, and using equation (51) reduces the GLRT of equation (54) to

$$\ell(\tilde{Z}) = \hat{\bar{m}}^T[B^TV^{-1}B]\,\hat{\bar{m}} = \hat{M}^TV^{-1}\hat{M}. \tag{55}$$

The estimate of the polynomial coefficients $\hat{\bar{m}}$ is also a Gaussian random variable with mean $\hat{\bar{m}}$ and covariance $[B^TV^{-1}B]$; consequently,

$$P(\bar{m}) \Rightarrow N(\hat{\bar{m}}, [B^TV^{-1}B]). \tag{56}$$

Because the covariance of M is symmetric positive definite, it can be factored as

$$B^T V^{-1} B = D^T D; \tag{57}$$

letting

$$r = D\,\hat{\bar{m}} \tag{58}$$

makes the GLRT become

$$\ell(\tilde{Z}) = r^T r \underset{H_1}{\overset{H_0}{\lessgtr}} \ln\lambda^2, \tag{59}$$

where

$$P(r) \Rightarrow N(D\,\hat{\bar{m}}, I)\ . \tag{60}$$

The test statistic is a sum of nonzero-mean, univariant, squared, Gaussian random variables; therefore, $P[\ell(\tilde{Z})]$ is a noncentral, chi-squared distribution with the number of degrees of freedom equal to the number of polynomial coefficients and a parameter of noncentrality of (references
4 - 6):

$$C = \hat{M}^T V^{-1} \hat{M} = \ell(\tilde{Z}). \tag{61}$$

Consider the example of special interest introduced previously. Substituting for $\hat{M}$ from equation (43)

$$\hat{M}_{i+1} = a_i \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ [(j-1)T]^i \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}; \tag{62}$$

and realizing that for this case $V = \sigma_n^2 I$ yields the test statistic

$$\ell_i(\tilde{Z}) = \frac{a_i^2}{\sigma_n^2 k_{2i}} \left(\sum_{j=1}^{K} (j-1)^i \tilde{z}_j \right)^2 \underset{H_1}{\overset{H_0}{\lessgtr}} \ln\lambda^2. \tag{63}$$

This form can be simplified so as to avoid evaluation of the noncentral, chi-squared density.

Because $\lambda > 1$, the right-hand side of the inequality in (63) is positive; hence, the square root of the test statistic can be used, resulting in

$$\left[\ell_i(\tilde{Z})\right]^{1/2} = \frac{a_i}{\sigma_n\sqrt{k_{2i}}}\left|\sum_{j=1}^{K} (j-1)^i\, \tilde{z}_j\right| \underset{H_1}{\overset{H_0}{\lessgtr}} \left[\ln\lambda^2\right]^{1/2}. \tag{64}$$

Let

$$r_i = \frac{a_i}{\sigma_n\sqrt{k_{2i}}} \sum_{j=1}^{K} (j-1)^i\, \tilde{z}_j; \tag{65}$$

then,

$$P(r_i|H_o) \Rightarrow N[0,1], \tag{66a}$$

$$P(r_i|H_1) \Rightarrow N[k_{ni},1]. \tag{66b}$$

Consequently, the test can be restated as

$$|r_i| \underset{H_1}{\overset{H_0}{\lessgtr}} [\ln\lambda^2]^{1/2} = \varepsilon_{GLRT}.$$

Figure 4 illustrates the normal density of r under the two hypotheses. Because the test statistic is the absolute value of r, the evaluation of P_F and P_D must include the areas under the density function for both positive and negative values of r that exceed the threshold; hence,

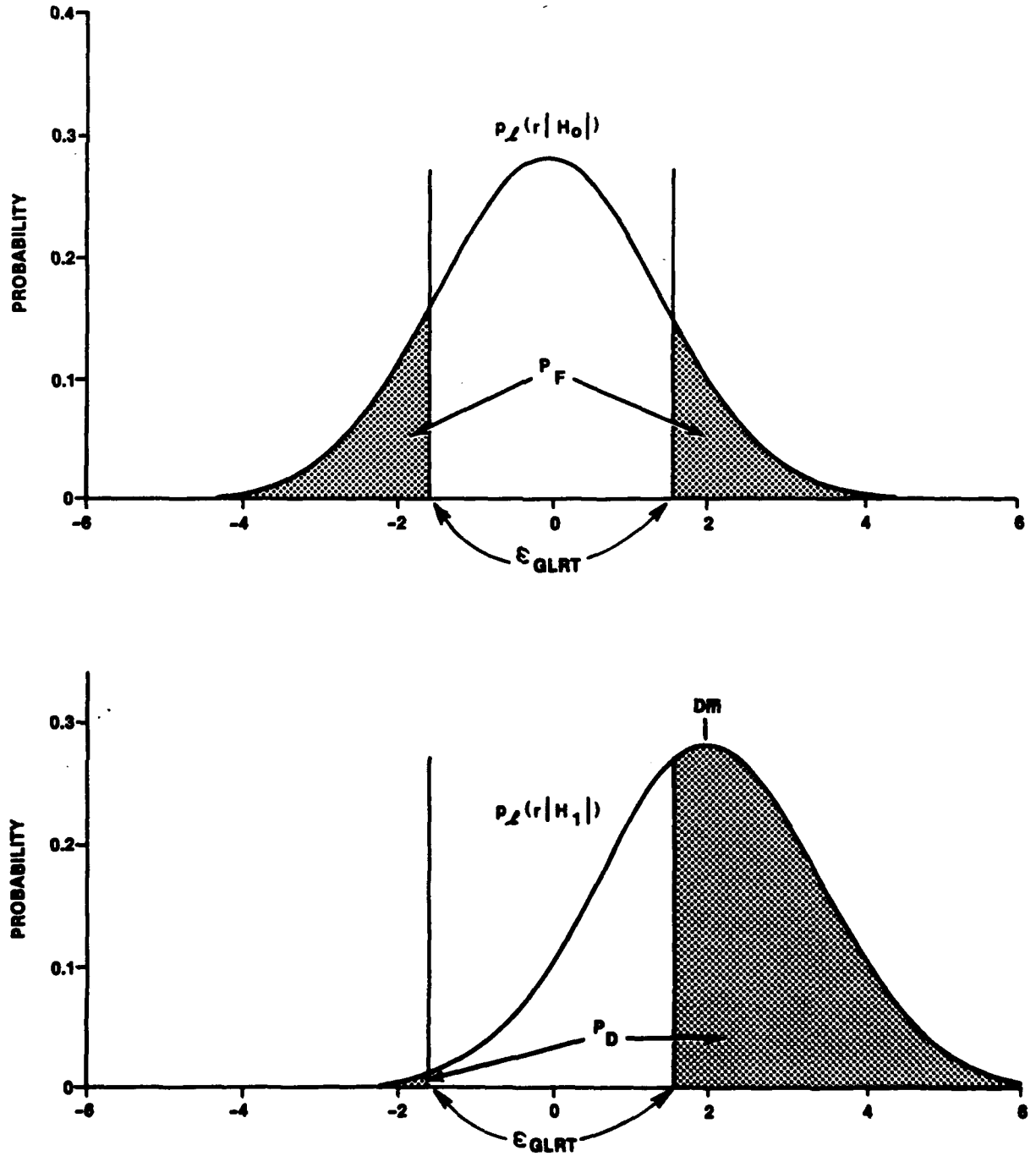


Figure 4. Binary GLRT

$$P_F = \int_{-\infty}^{-\varepsilon_{GLRT}} P(r_i|H_0)dr_i + \int_{\varepsilon_{GLRT}}^{-\infty} P(r_i|H_0)dr_i, \tag{67a}$$

$$P_F = 2(1 - \text{erf}\,[\varepsilon_{GLRT}]), \tag{67b}$$

and

$$P_D = \int_{-\infty}^{-\varepsilon_{GLRT}} P(r_i|H_1)dr_i + \int_{\varepsilon_{GLRT}}^{-\infty} P(r_i|H_1)dr_i, \tag{68a}$$

$$P_D = \text{erf}\,[-\varepsilon_{GLRT} - k_{ni}] + 1 - \text{erf}\,[\varepsilon_{GLRT} - k_{ni}]. \tag{68b}$$

Note that the shaded areas of the density models in figure 4 are probability masses of the corresponding single-degree-of-freedom, chi-squared and noncentral, chi-squared random variables of r_i^2 given H_0 and r_i^2 given H_1, respectively. For a given false alarm probability the LRT detection probability is greater than the GLRT detection probability. A plot of the ROC for the GLRT is given in figure 5. Although not apparent in the figure, the LRT outperforms the GLRT by a maximum of approximately 14 percent in probability of detection.

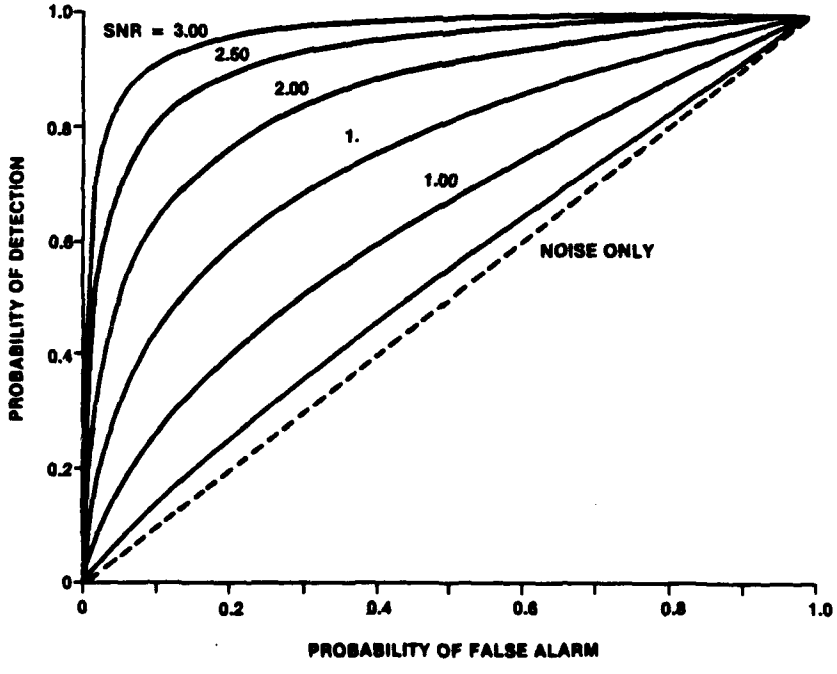


Figure 5. GLRT Receiver Operating Characteristic

3.3 ARTIFICIAL NEURAL NETWORK (ANN) HYPOTHESIS TEST

Another approach to hypothesis testing is to employ an artificial neural network (ANN) that emulates the hypothesis testing algorithm. An ANN is an interconnected network of nodes and branches. The branches are weighted and serve as both inputs and outputs for the nodes. The nodes sum the weighted inputs and provide a single output that is a function (usually nonlinear) of the sum of the inputs. Typical architectures use layers of nodes with the output of each node of one layer providing an input to every node of the next layer. Figure 6 illustrates a single node with six inputs and a sigmoid output function. A typical feedforward architecture with two hidden layers is illustrated in figure 7. Note that the output layer need not be a single node, and that in general the number of hidden layers, the number of nodes in each layer, as well as the types of function(s) to be used are part of the "art" of designing an ANN. Given a set of input n-tuples and corresponding p-tuple outputs, the weights (w_i) and biases (b_j) that provide the best fit to the input-output map can be determined. The process of finding the weights and biases is termed "training," and the data set on which training is performed is called the "training set."

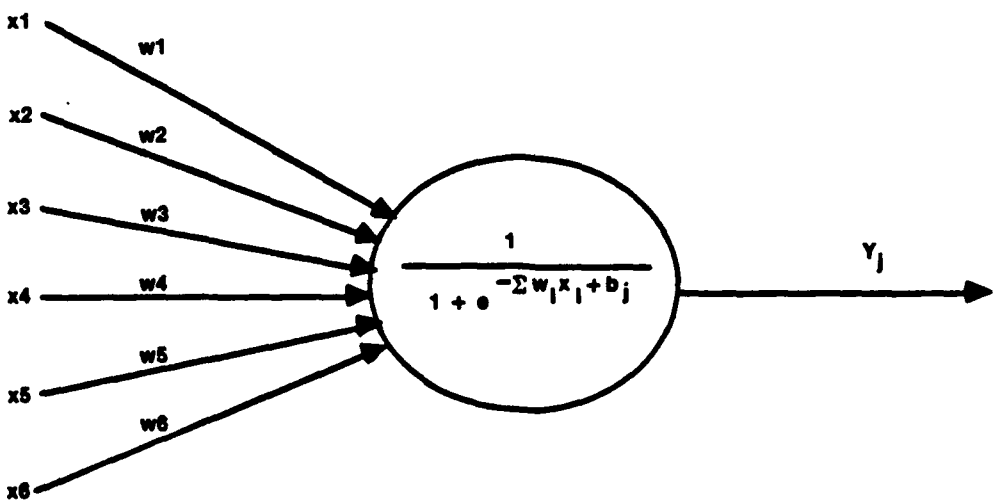


Figure 6. Typical Single Neuron

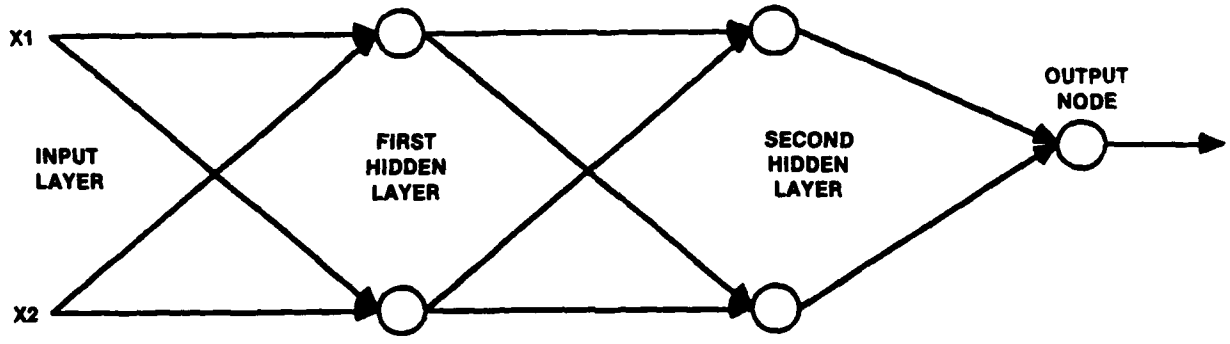


Figure 7. Typical Feedforward Network with Two Hidden Layers

3.4 TRAINING THE ARTIFICIAL NEURAL NETWORK

The function used in the nodes will determine the procedure needed to train the neural network. Here a sigmoid, which is a smooth, continuous, nonlinear function is used. These characteristics allow standard nonlinear search techniques to be employed in training.

Training also entails having an input/output relationship for the ANN to emulate. In this case the network will be required to produce a function of the test statistic of equation (48), $\ell_i(\tilde{Z})$, when $\tilde{Z}$ is given as an input. The sigmoid function defined in figure 6 has an output in the range [0,1] as shown in figure 8.

The function of the test statistic is defined as follows. For a feature made with a positive coefficient, the required output function $y_i(\tilde{Z})$ is defined with the sigmoid function as $\{1.0 + \exp[-\alpha\ell_i(\tilde{Z})]\}^{-1}$, where α is chosen so that the output achieved the specific value of a 2-percent probability of false alarm at the output levels of 0.5 ± 0.1. Similarly, if the coefficient is negative, $y_i(\tilde{Z})$ is defined as $\{1.0 + \exp[\alpha\ell_i(\tilde{Z})]\}^{-1}$. Note that these two functions give a continuous output from zero to one as the feature parameter a_i goes from minus infinity to plus infinity, that $\ell_i(\tilde{Z})$ can be readily computed by the inverse mapping of the output function y_i, and the polarity of the feature is indicated.

Fifteen data points were determined to be an appropriate tradeoff between a longer integration time for better feature resolution and detection, and a shorter record length to limit the number of features likely to occur in a given sequence. Longer sequences can be accommodated by sliding the data window.

The software used to train the network also required the inputs to be non-negative. To overcome this and for ease of training the network to recognize features, the inputs to the network were scaled and shifted as

$$\tilde{z}_n(i) = \tilde{z}(i) \quad \frac{0.0125}{\sigma_n} + 0.5, \tag{69}$$

where the scale factor of 0.0125 was arbitrarily selected to yield a dynamic range of 80 measurement noise standard deviations for the combined waveforms.

The ANN was trained as follows. Given residual vectors and their associated output functions, compare the network's outputs to the required output functions and adjust the weights and biases to minimize the error. This procedure is illustrated in figure 9. While the training rule used to adjust the weights and biases can be any nonlinear search technique, a quasi-Newton technique was employed here.

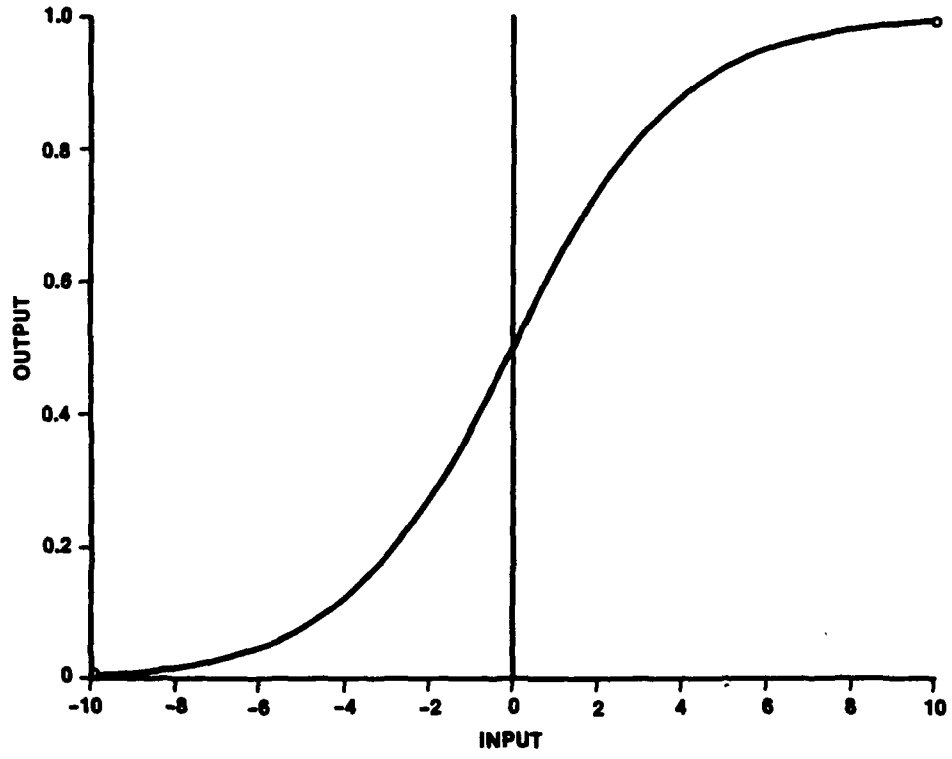


Figure 8. Illustration of the Sigmoid Function

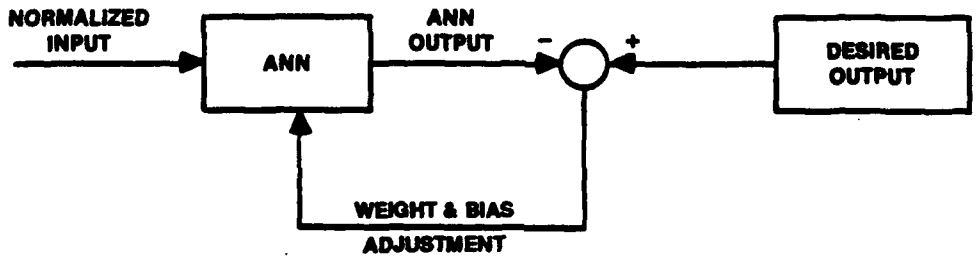


Figure 9. Illustration of Network Training

4. MULTIPLE HYPOTHESIS TEST

In section 3 the problem of detecting a signal with known structure in the presence of noise was considered. Because that test considers one of two possible outcomes (either the signal is present or it is not), it is often called a binary hypothesis test. This section discusses the multiple hypothesis tests.

4.1 LIKELIHOOD RATIO TEST

The anomalous feature model was defined by equations (12) and (13) as a $(q - 1)^{th}$ order polynomial. In the binary hypothesis test of section 3, the signal detection problem was considered. In that case the signal structure was known; that is, polynomial order was known as well as which coefficients were nonzero. Here, only the maximum model order is known, and a test is devised to determine which polynomial coefficients are nonzero. This is a standard problem in regression analysis and the development presented here has been adapted from references 6 and 7.

The case considered here is a maximum polynomial order of one for the anomalous feature; that is,

$$\bar{m} = [a_0, a_1]^T. \tag{70}$$

The hypotheses are

H_0: $[a_0 = 0, a_1 = 0]$ no anomaly present - noise only,
H_1: $[a_0 \neq 0, a_1 = 0]$ jump anomaly only,
H_2: $[a_0 = 0, a_1 \neq 0]$ drift anomaly only,
H_3: $[a_0 \neq 0, a_1 \neq 0]$ jump and drift anomalies.

The development is easily extended to arbitrary model order by noting that the set of polynomial coefficients results in 2^q hypotheses.

Consider the special case of uncorrelated error in the predicted residual sequence. This corresponds to the case of negligible estimation error. In this case the model for the predicted residual of equation (13) is

$$\tilde{Z} = N_K + M. \tag{71}$$

Because the measurement noise vector N_K is an independent, zero-mean, Gaussian random vector, it follows that

$$P(\tilde{Z}) \Rightarrow N[M,\sigma_n^2 I].$$

The joint density function of the predicted residuals, conditioned on the hypothesis H_i that the i^{th} anomalous feature set is present in the sequence (i.e., jump-only, drift-only, jump-drift, etc.), is

$$P(\tilde{Z}|H_i) \Rightarrow N[M_i,\sigma_n^2 I],$$

where M_i is the corresponding mean vector induced by the anomaly of the i^{th} hypothesis. The likelihood ratio test between any two hypotheses H_i and H_j is

$$LRT(\tilde{Z}) = \frac{P(\tilde{Z}|H_i)}{P(\tilde{Z}|H_j)} \underset{\bar{H}_i}{\overset{\bar{H}_j}{\gtrless}} \lambda. \tag{72}$$

Here, if the LRT exceeds the threshold λ, then hypothesis H_j is rejected. Likewise, if the LRT is below the threshold, then hypothesis H_i is rejected.

The case where the measurement noise standard deviation σ_n is known and the anomaly $M = [m_1, ..., m_N]^T$ is unknown is considered first. Substituting for $P(\tilde{Z}|H)$ for the general case results in

$$LRT(\tilde{Z}) = \frac{[(2\pi)^{K/2} \det |V|^{1/2}]^{-1} \exp[-\frac{1}{2}(\tilde{Z} - M_i)^T V^{-1}(\tilde{Z} - M_i)]}{[(2\pi)^{K/2} \det |V|^{1/2}]^{-1} \exp[-\frac{1}{2}(\tilde{Z} - M_j)^T V^{-1}(\tilde{Z} - M_j)]} \underset{\bar{H}_i}{\overset{\bar{H}_j}{\gtrless}} \lambda. \tag{73}$$

Taking the natural logarithm yields

$$\ln[LRT(\tilde{Z})] = -\frac{1}{2}\|\tilde{Z} - M_i\|^2_{V^{-1}} + \frac{1}{2}\|\tilde{Z} - M_j\|^2_{V^{-1}} \underset{\bar{H}_i}{\overset{\bar{H}_j}{\gtrless}} \ln\lambda. \tag{74}$$

For the case of interest, $V = \sigma_n^2 I$ and, because M is unknown, the estimated $\hat{M}$ of equations (51) and (52) is used, resulting in

$$\hat{M} = B\,\hat{\underline{m}}, \tag{75a}$$

$$\hat{\underline{m}} = [B^T B]^{-1} B^T Z, \tag{75b}$$

which upon substituting in (74) yields the GLRT:

$$\ln[GLRT(\tilde{Z})] = \frac{\|\tilde{Z} - \hat{M}_j\|^2}{\sigma_n^2} - \frac{\|\tilde{Z} - \hat{M}_i\|^2}{\sigma_n^2} \underset{\bar{H}_i}{\overset{\bar{H}_j}{\gtrless}} \ln\lambda^2 , \tag{76a}$$

or

$$\sum_{k=1}^{K} \frac{(\tilde{z}_k - \hat{m}_{jk})^2}{\sigma_n^2} - \sum_{k=1}^{K} \frac{(\tilde{z}_i - \hat{m}_{ik})^2}{\sigma_n^2} \underset{\bar{H}_i}{\overset{\bar{H}_j}{\gtrless}} \ln\lambda^2 , \tag{76b}$$

where $\hat{m}_{jk}$ is the k^{th} component of the vector $\hat{M}_j$. Define the test statistic for the multiple hypothesis generalized likelihood ratio test as

$$\ell_m(\tilde{Z}) = \ln [GLRT(\tilde{Z})]. \tag{76c}$$

To obtain quantifiable results for the GLRT, it is necessary to know the probability density function (PDF) of the test statistic ℓ_m. The remainder of this section is devoted to the development of the test statistic's PDF.

From equation (13), the mean vector M can be decomposed into a linear combination of columns of the matrix B; that is,

$$M = a_0 B_1 + a_1 B_2 + \ldots + a_{q-1} B_q, \tag{77a}$$

where B_i are the columns of B:

$$B = [B_1 | \ldots | B_q] , \tag{77b}$$

and

$$\bar{m} = [a_0, ..., a_{q-1}]^T. \tag{77c}$$

The vectors $\{B_i\}_{i=1}^{q}$ are independent and therefore form the basis for a q-dimensional subspace of the K-dimensional space of the predicted residual vector. Let the j^{th} hypothesis be the anomaly modeled by the $q\text{-}1^{th}$ order polynomial with all coefficients present (H_3 for the case of interest). The other hypotheses are developed as constraints on the coefficients; that is,

$$C\bar{m} = 0. \tag{78}$$

For the case of interest here, the hypothesis H_i results in the following constraint matrices C:

$$H_0: C_0 = I_{2x2}, \qquad H_1: C_1 = [0,1], \qquad H_2: C_2 = [1,0],$$

where I_{2x2} is the 2x2 identity matrix. When hypothesis $H_i (i \neq j)$ is true, p-constraints are imposed on the coefficients a_i that can be used to reduce the number of independent terms in the decomposition of M in equation (77). Consequently, if $H_i = H_1$, then $a_1 = 0$, and

$$M_1 = a_0 B_1, \tag{79}$$

where the rank of C equals p and the space spanned by the remaining columns of B has been reduced by one. In general, if p constraints are imposed by the i^{th} hypothesis, then

$$M_i = a_0' \beta_1 + ... + a_{q-p-1}' \beta_{q-p}, \tag{80}$$

where a_i' are the surviving coefficients and β_i are the basis vectors of (q - p)-dimensional subspace spanned by M under H_i.

Let the set of Kx1 vectors $\{r_i\}_{i=1}^{K}$ be an orthonormal basis for the K-dimensional space of the predicted residuals. Specifically, let the first (q - p) vectors $\{r_i\}_{i=1}^{q-p}$ be the basis for the space spanned by the constrained columns of B, i.e., $\{\beta_i\}_{i=1}^{p-q}$ and, further, let the set of vectors $\{r_i\}_{i=1}^{q}$ span the column space of B; then,

$$\tilde{Z} = R\underline{\alpha}, \tag{81a}$$

and

$$M = R\underline{\gamma}, \tag{81b}$$

where

$$R = [r| \ldots |r_K], \tag{81c}$$

and $\underline{\alpha}$ and $\underline{\gamma}$ are Kx1 vectors of the coefficients of $\tilde{Z}$ and M, respectively, with respect to the orthonormal basis vectors $\{r_i\}_{i=1}^{K}$ and

$$\underline{\alpha} = [\alpha_1 \ldots \alpha_K]^T, \tag{81d}$$

and

$$\underline{\gamma} = [\gamma_1, \ldots, \gamma_{q-p}, 0, \ldots, 0]^T. \tag{81e}$$

Substituting equation (81) into the two forms of equation (74), under the condition that test hypothesis H_j is true, results in

$$\|\tilde{Z} - M_j\|^2 = \sum_{k=1}^{q} (\alpha_k - \gamma_k)^2 + \sum_{k=q+1}^{K} \alpha_k{}^2, \tag{82a}$$

and

$$\|\tilde{Z} - M_i\|^2 = \sum_{k=1}^{q-p} (\alpha_k - \gamma_k)^2 + \sum_{k=q-p+1}^{K} \alpha_k{}^2. \tag{82b}$$

For the GLRT, M is not available and must be estimated. The estimate is given by equation (75), which for the case of interest here ($V = \sigma^2 I$) is equivalently the least-squares estimate or the maximum likelihood estimate. Because the estimate of M must minimize the sum of the squared error in the fit of the mean M to the predicted residuals $\tilde{Z}$, it follows that M must minimize equation (82). From the right-hand side of equation (82), the least-squares estimate of $\hat{M} = R\hat{\gamma}$ is

$$H_i: \quad \hat{\gamma}_k = \alpha_k; \qquad k = 1, \ldots, q - p, \tag{83a}$$

$$H_j: \quad \hat{\gamma}_k = \alpha_k; \qquad k = 1, \ldots, q. \tag{83b}$$

Consequently, equation (82) reduces to

$$\| \tilde{Z} - \hat{M}_j \|^2 = \sum_{k=q+1}^{K} \alpha_k{}^2, \tag{84a}$$

and

$$\| \tilde{Z} - \hat{M}_i \|^2 = \sum_{k=q-p+1}^{K} \alpha_k{}^2. \tag{84b}$$

Substituting equation (84) into the GLRT of equation (76) and reducing yields

$$\ell_m(\tilde{Z}) = \frac{1}{\sigma_n^2}\left(\sum_{k=q-p+1}^{q} \alpha_k{}^2 \right) \underset{H_j}{\overset{H_i}{\gtrless}} \ln \lambda^{-2}. \tag{85}$$

The predicted residual $\tilde{Z}$ is a Gaussian random vector with mean M_i, covariance $\sigma_n^2 I$, and PDF given by

$$P(\tilde{Z}) \Rightarrow N(M_i, \sigma_n^2 I). \tag{86}$$

From equation (82), the PDF of the random vector $\underline{\alpha}$ is

$$P(\underline{\alpha}) = [\det|R^{-1}|]^{-1} P(R^{-1}\tilde{Z}). \tag{87}$$

Because R is a linear orthogonal transformation ($R^{-1} = R^T$ and $\det|R| = 1$), it follows that the random vector $\underline{\alpha}$ is also Gaussian with PDF

$$P(\underline{\alpha}) \Rightarrow N(\underline{\gamma}, \sigma_n^2 I). \tag{88}$$

Consequently, the GLRT statistic ℓ_m is the sum of the square of p, zero-mean, unit-variance, independent, Gaussian random variables and is therefore a chi-squared random variable with p degrees (rank of the constraint matrix) of freedom. Recall that this result was obtained under the condition that H_i is true. Therefore, the test statistics ℓ_m will exceed the threshold and reject H_i when H_i is true with a probability that is given by the chi-squared distribution with p degrees of freedom. Similarly, if H_j is true, then at least some of the random variables $\{\alpha_k: k = q - p, ..., q\}$ are not zero mean, and the test statistic becomes a noncentral, chi-squared random variable with p degrees of freedom. Under the conditions of H_j, the test statistic increases, thereby increasing the probability that the threshold will be exceeded and H_i rejected.

The GLRT statistic can be represented in a form that is more convenient to use. Noting from equations (81) and (82) that

$$\hat{M}_j = \sum_{k=1}^{q} \alpha_k r_k, \tag{89a}$$

and

$$\hat{M}_i = \sum_{k=1}^{q-p} \alpha_k r_k, \tag{89b}$$

it follows that the difference in the estimates of the means under hypotheses j and i is

$$[\hat{M}_j - \hat{M}_i] = \sum_{k=q-p+1}^{q} \alpha_k r_k, \tag{89c}$$

and the squared norm of the difference of the means is

$$\| \hat{M}_j - \hat{M}_i \|^2 = \sum_{k=q-p+1}^{q} \alpha_k{}^2 \ . \tag{90}$$

Note, however, that the right-hand side of (90) is the unnormalized GLRT statistic of equation 85; hence,

$$\| \tilde{Z} - \hat{M}_j \|^2 - \| \tilde{Z} - \hat{M}_i \|^2 = \| \hat{M}_j - \hat{M}_i \|^2, \tag{91a}$$

or

$$\ell_m(\tilde{Z}) = \frac{\| \hat{M}_j - \hat{M}_i \|^2}{\sigma_n^2} \underset{H_j}{\overset{H_i}{\gtrless}} \ln\lambda^{-2}. \tag{91b}$$

That is, the GLRT statistic reduces to the squared norm of the distance between the two mean vectors estimated under the j^{th} and i^{th} hypotheses. Given that H_j is the full order model, then ℓ_m is central chi-squared with p degrees of freedom when H_i is true and is non-central chi-squared with p degrees of freedom when H_j is true.

For the specific case of j = 3, the tests of interest are

$$T_2: \frac{\|\hat{M}_3 - \hat{M}_2\|^2}{\sigma_n^2} \underset{H_3}{\overset{H_2}{\gtrless}} \varepsilon_2, \tag{92a}$$

$$T_1: \frac{\|\hat{M}_3 - \hat{M}_1\|^2}{\sigma_n^2} \underset{H_3}{\overset{H_1}{\gtrless}} \varepsilon_1, \tag{92b}$$

$$T_0: \frac{\|\hat{M}_3\|^2}{\sigma_n^2} \underset{H_3}{\overset{H_0}{\gtrless}} \varepsilon_0. \tag{92c}$$

Tests T_i evaluate the statistical significance of the presence of each coefficient or feature in the polynomial indicative of anomalies in the predicted residual sequence. For T_2 the test statistic is the norm-squared value in units of noise standard deviation of the difference between the estimates of the mean obtained under the hypothesis that jump and drift features are present ($M_3 \Rightarrow a_0 \neq 0, a_1 \neq 0$) and the hypothesis that only drift is present ($M_2 \Rightarrow a_0 = 0, a_1 \neq 0$). Thus, T_2 is a test of the significance of the nonzero estimate of the jump coefficient a_0. If the test statistic is below threshold, the hypothesis that jump and drift are present is rejected in favor of the hypothesis that only a drift feature is present, i.e., no significant jump feature is evident. Conversely, if the test statistic exceeds the threshold, the hypothesis that only a drift is present (H_2) is rejected in favor of the hypothesis that both jump and drift features (H_3) are evident in the sequence.

When H_2 is true, which implies a_0 is zero, the test statistic is central chi-squared with one degree of freedom. Consequently, the threshold ε_2 can be established to give a desired probability of false detection of the jump feature. When H_3 is true, a jump is present and the test statistic is

noncentral chi-squared with one degree of freedom. Although the value of the jump coefficient a_0 is unknown, the parameter of noncentrality could be computed for some value of a_0 and the threshold ε_2 set to achieve a specified probability of jump detection at a given amplitude. However, this approach was not used. Notice that rejection of hypothesis H_3 by test T_2 does not imply that the drift hypothesis is accepted. The only conclusion that can be drawn from test T_2 is whether or not jump is present. Similar reasoning applies for test T_1, which determines if drift is present. The T_3 test determines if the combined hypothesis of jump plus drift is to be accepted/rejected over the null hypothesis of no anomaly (H_0). This test statistic is central chi-squared with two degrees of freedom when H_0 is true. (Test T_0 is redundant with tests T_1 and T_2 and is not used.)

To complete the hypothesis test, a logic tree is constructed from the possible outcomes listed in table 1.

Table 1. Multiple Hypothesis Test

T_2 / T_1	$(a_0 \neq 0)$ $\bar{H}_2$	$(a_0 = 0)$ $\bar{H}_3$
$(a_1 \neq 0)$ $\bar{H}_1$	H_3	H_2
$(a_1 = 0)$ $\bar{H}_3$	H_1	H_0

If the result of T_1 indicates the presence of drift (reject $H_1 \Rightarrow a_1 \neq 0$) and that of T_2 indicates the absence of a jump (reject $H_3 \Rightarrow a_0 = 0$), the logical conclusion is that the predicted residuals exhibit a drift-only feature (H_2).

4.2 MODIFIED NEYMAN-PEARSON TEST

Multiple feature detection can also be performed as a sequence of independent binary hypothesis tests with the aid of an *ad hoc* modification to the predicted residuals. With this approach, an estimate for the feature, $\hat{\bar{m}}$ is obtained under the hypothesis that all the features are present, H_3 in this case. New data sequences are formed that contain a single feature by subtracting the other features; that is,

$$\tilde{Z}_1 = \tilde{Z} - \hat{M}_2, \tag{93a}$$

$$\tilde{Z}_2 = \tilde{Z} - \hat{M}_1, \tag{93b}$$

where

$$\hat{M}_1 = BE_1\hat{\bar{m}}, \tag{93c}$$

$$\hat{M}_2 = BE_2\hat{\bar{m}}, \tag{93d}$$

and

$$E_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \tag{93e}$$

$$E_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \tag{93f}$$

Here, $\tilde{Z}_1$ is a pseudo-predicted residual sequence that, in principle, has had all the features removed except jump. Similarly, the sequence $\tilde{Z}_2$ is devoid of all features except drift. Previously, it was shown that the feature model of H_3 will give a better fit to $\tilde{Z}$ than the other feature models containing fewer coefficients. Thus, in some sense, $\tilde{Z}_1$ and $\tilde{Z}_2$ represent separate observations of the individual features, which allows for independent binary tests on each feature as in equation (31). However, when using $\tilde{Z}_i$ and $\hat{M}_i$ in equation (31) it must be noted that the results of each binary test are obtained assuming the presence of all the other features. That is, jump exists (does not exist) given that a drift is present, etc. Recall that equation (31) is the test statistic for the binary hypothesis general likelihood ratio test. Because the LRT is derived from the Neyman-Pearson test criteria, this technique is called the modified Neyman-Pearson (MNP) test. Note that only one set of coefficient estimates $\hat{\bar{m}}$ is required, which is a significant computational savings over the chi-squared test. However, the probability density function for the MNP test statistic was not determined. Consequently, the thresholds needed to give the required probabilities of false alarm and detection are not determined.

4.3 ARTIFICIAL NEURAL NETWORK

The same ANN previously described for the binary hypothesis test was used for the detection of multiple features. The only difference was that the multiple hypothesis ANN was constructed with two output nodes versus the single output node for the binary ANN. The network was trained to emulate the likelihood functions produced by the MNP test. The MNP results were selected for training because of the MNP algorithm's computational efficiency and, as will be discussed later, the fact that its performance was nearly identical to that of the chi-squared test.

5. EXPERIMENTAL RESULTS AND DISCUSSION

The first phase in the process of evaluating the neural network's performance involved training an ANN to emulate the LRT binary hypothesis test of detecting a jump in the presence of noise. (The training method was described earlier in section 3.4.) In this phase, a network with a single output was used. ROC curves were produced using synthetic data sequences generated by adding Gaussian white noise, with zero mean and unit variance, to jumps with various amplitudes, providing a range of signal-to-noise ratio (SNR) values, and comparing the resulting output of the ANN with a range of thresholds. Here, SNR is defined as

$$SNR = \sqrt{M^T V^{-1} M},$$

where V is the covariance matrix $\sigma_n^2 I$. The results were accumulated for 10,000 Monte Carlo trials for each of the SNR levels shown in figure 3, which is a plot of $P(H_1|H_1, \varepsilon)$ for varying ε, and specific jump amplitudes as a parameter. It can be seen (figures 10 and 11) that the ANN matched the theoretical predictions of a GLRT. These results show that it is possible to train an ANN to efficiently emulate a Neyman-Pearson test procedure.

Experiments were then conducted to compare the ANN's performance as a feature detector and discriminator against the χSQ and MNP techniques when multiple features are present. Here, a synthetic data sequence was generated as previously described, except that the signal levels were defined by

$$SL(a_0) = (a_0\sqrt{K})/\sigma_n ,$$

and

$$SL(a_1) = a1\,\frac{\sqrt{2K^3 - 3K^2 + K}}{6\sigma_n} .$$

The results were accumulated as follows. Figures 12a and 12b show the overall probability of detecting a jump feature given that a jump is present (regardless of the presence of drift) versus the overall probability of detecting a jump given that no jump is present (again regardless of drift) for the χSQ test and the ANN, respectively. That is, $[P(H_1|H_1,\varepsilon) + P(H_1|H_3,\varepsilon)]$ is plotted versus $[P(H_1|H_0,\varepsilon) + P(H_1|H_2,\varepsilon)]$. As can be seen from the figures, the ANN somewhat outperforms the χSQ test for low- to mid-range false alarm probabilities but is subsequently outperformed at higher false alarm probabilities. The results for the MNP were nearly identical to those for the χSQ test and are not presented. A similar relationship (not shown) was obtained for detection of drift by the three techniques with similar results.

Figures 13a-d and 14a-d show the components that make up the overall curves of figures 12a and 12b, respectively. Figures 13a and 14a show (for the χSQ test and ANN, respectively)

the probability of detecting a jump when only drift is present versus the probability of detecting a jump when only noise is present (a false alarm), i.e., $P(H_1|H_2,\varepsilon)$ is plotted versus $P(H_1|H_0,\varepsilon)$. Figures 13b and 14b show the probability of detecting a jump given that a jump occurred versus the false alarm probability, i.e., $P(H_1|H_1,\varepsilon)$ vs $P(H_1|H_0,\varepsilon)$. Figures 13c and 14c show the probability of detecting a jump in the presence of an interfering drift with the same polarity versus the false alarm probability, i.e., $P(H_1|H_3,\varepsilon)$ vs $P(H_1|H_0,\varepsilon)$. Finally, figures 13d and 14d are constructed in the same manner as 13c and 14c except that the features have opposite polarity.

Figures 13a-d clearly show that the χSQ test for jump is insensitive to the presence of an interfering drift. However, examination of figures 14a-d shows that the ANN has an unwanted sensitivity to the interfering signal. This is especially evident in a comparison of figures 14c and 14d. Here, a drift with the same polarity aids detection of a jump, while a drift with opposite polarity inhibits detection. Similar results were obtained for drift and, once more, the MNP results were nearly identical to the χSQ results. The difference in sensitivity to interfering signals between the ANN and the other two techniques can be seen by comparing figures 13a and 14a.

When one views the overall curves of figure 12, the ANN appears to produce somewhat better performance than the χSQ test. This is the result of two phenomena. The first is that there are more detections of the desired signal when it is present due to the interfering signal (see figures 13b and 14b), and the second is a thresholding effect of the sensitivity to the interfering signal at low false alarm probabilities (see figure 14a). For these reasons, care was taken in comparing these techniques. This view is consistent with the concerns brought out in Lau and Widrow (reference 8). It is felt that, because the MNP (which the ANN was trained to emulate) does not exhibit this sensitivity to an interfering signal, and the ANN was successfully trained to emulate the GLRT in the binary hypothesis test, the sensitivity of the ANN can be overcome by a better selection of the training pattern set and/or a different network architecture. That is, the ANN has yet to fully learn this input/output relationship.

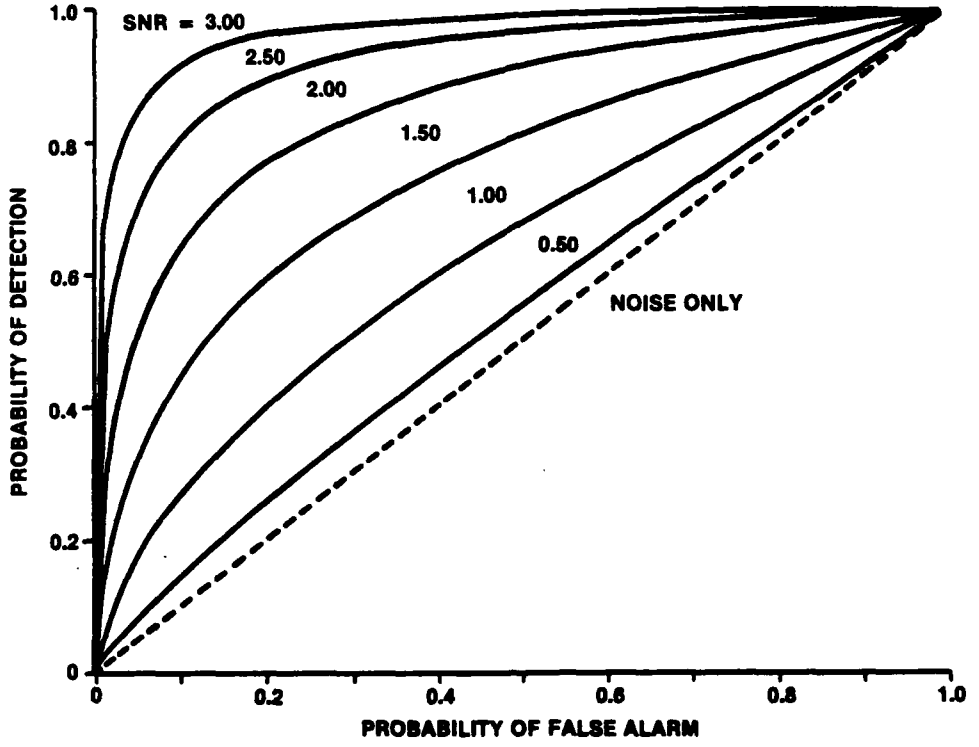


Figure 10. ANN Receiver Operating Characteristic (Jump Only)

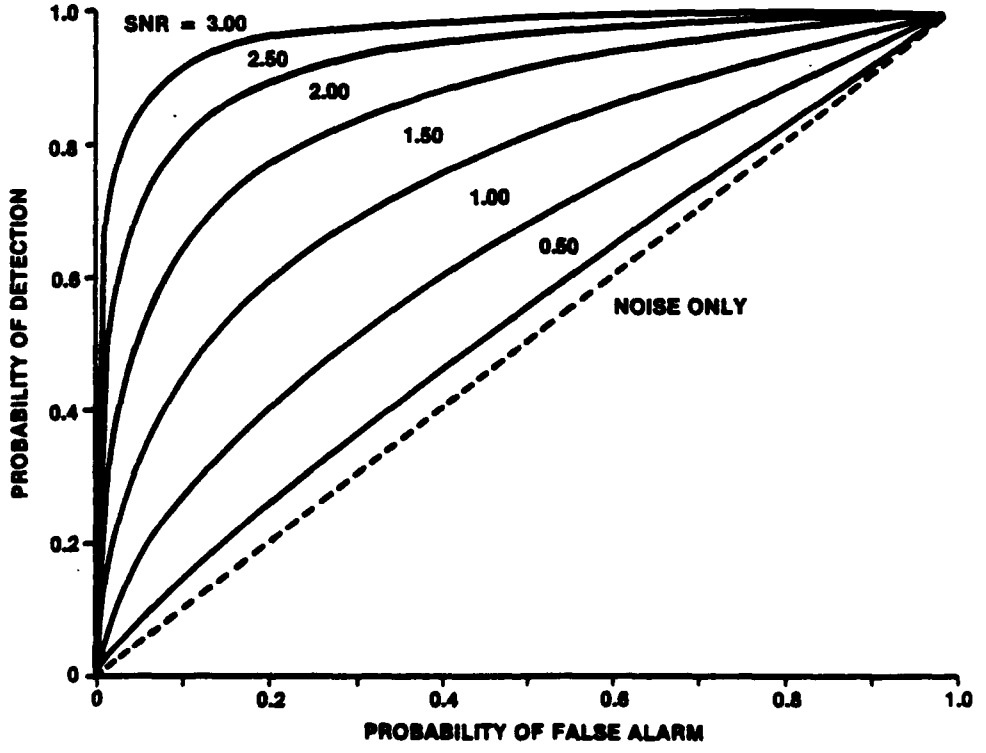


Figure 11. Theoretical GLRT Receiver Operating Characteristic (Jump Only)

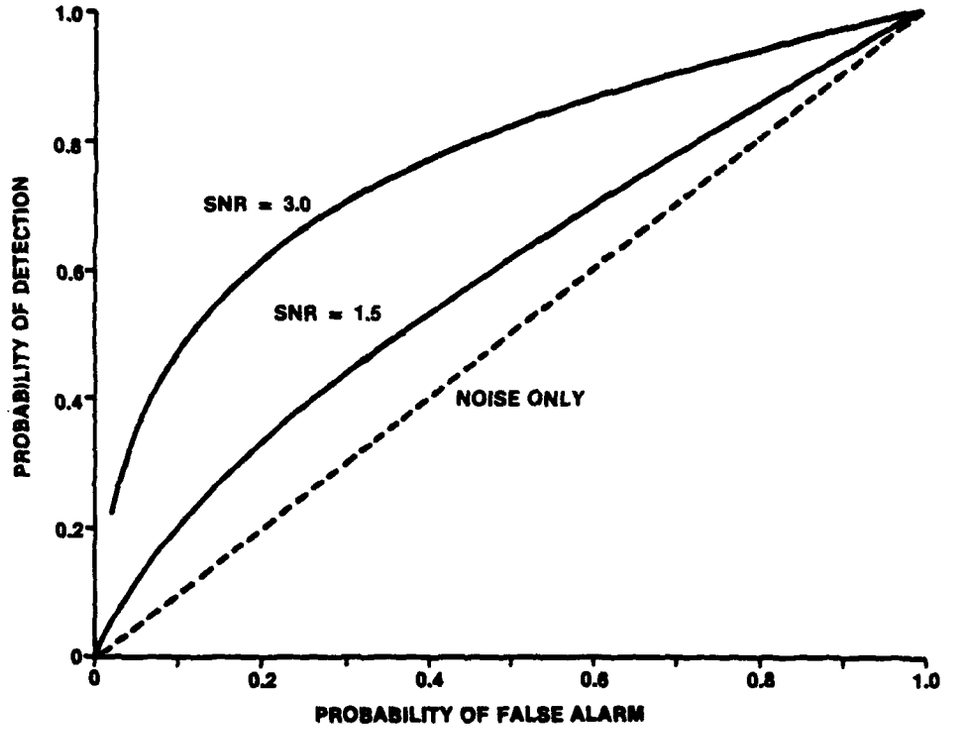


Figure 12a. χSQ Receiver Operating Characteristic (Jump)

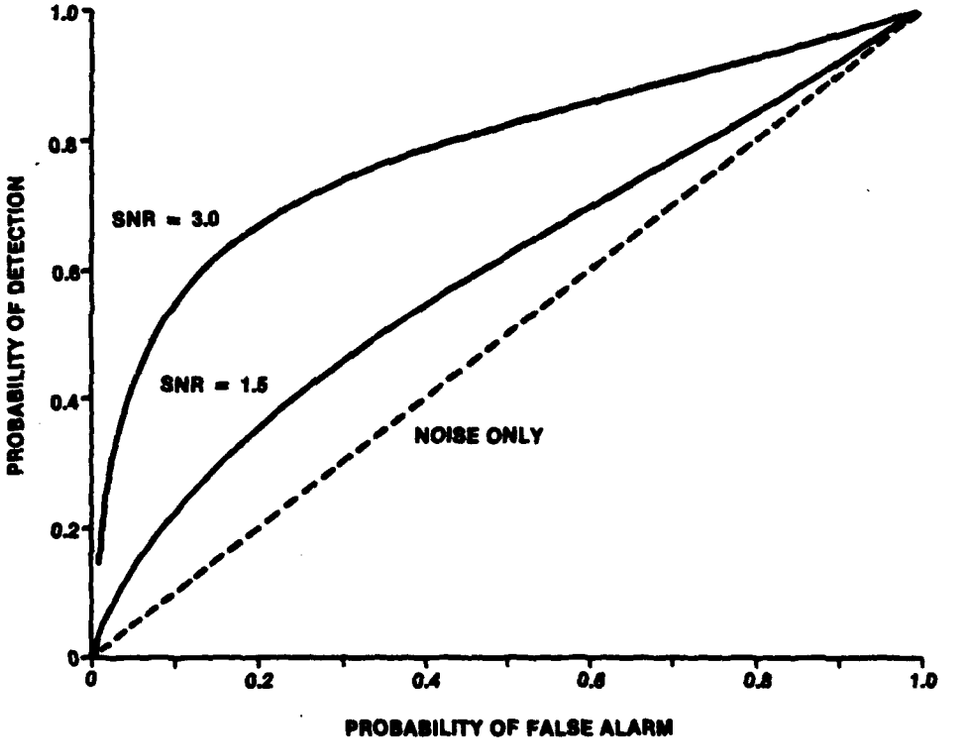


Figure 12b. ANN Receiver Operating Characteristic (Jump)

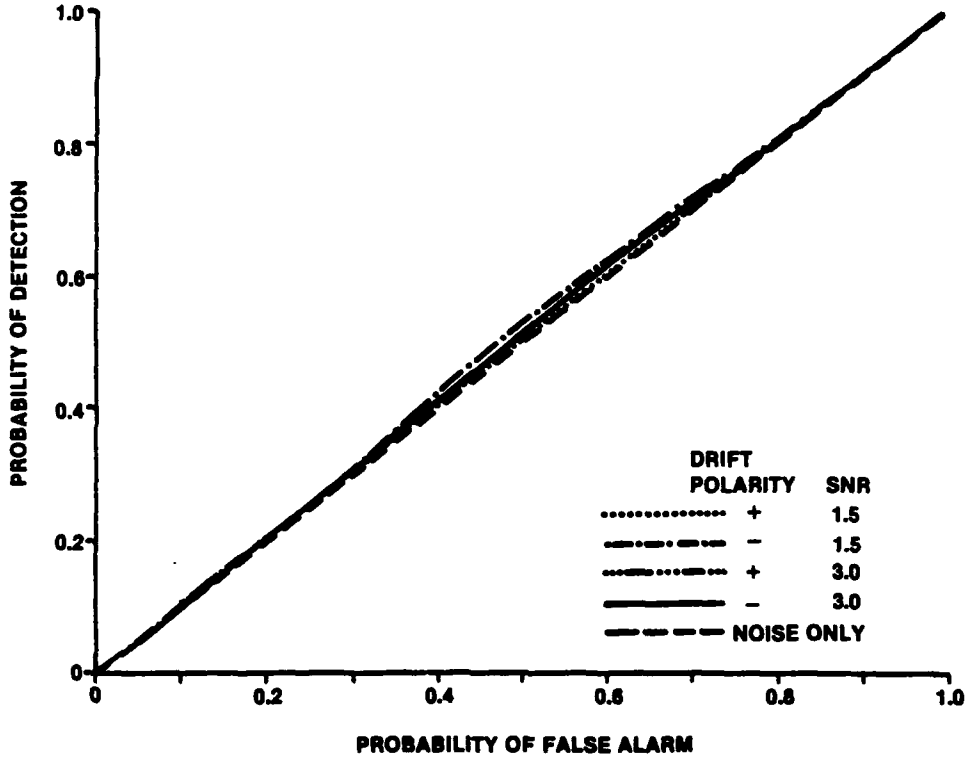


Figure 13a. χSQ Jump Output Given Drift Only

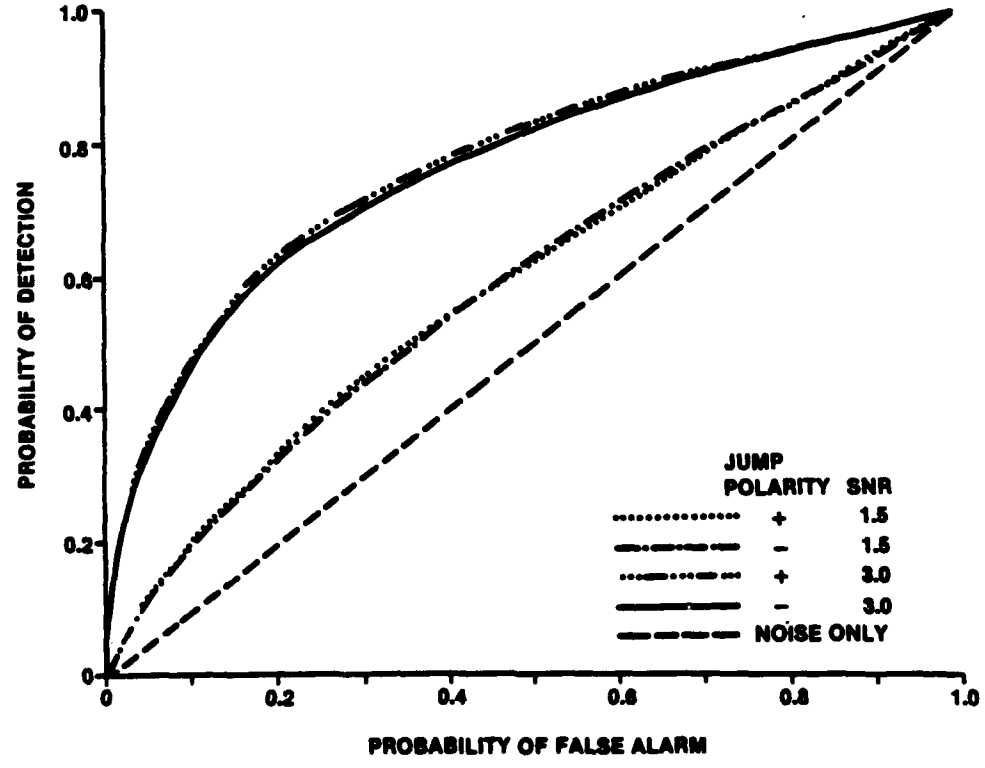


Figure 13b. χSQ Jump Output Given Jump Only

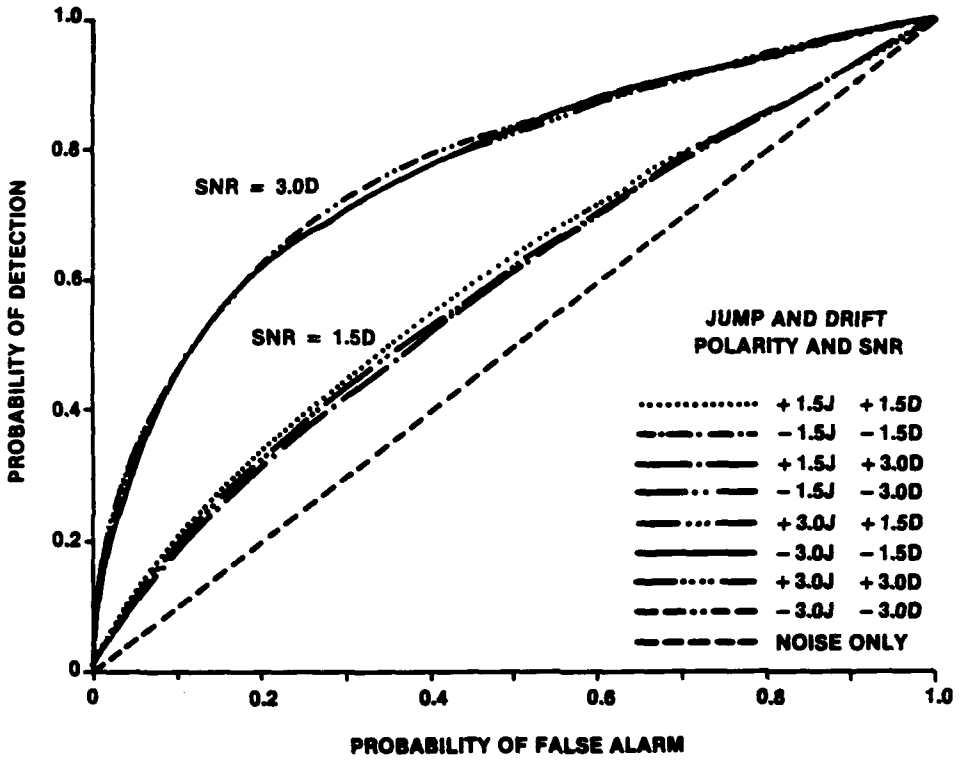


Figure 13c. χSQ Jump Output Given Drift with Same Polarity

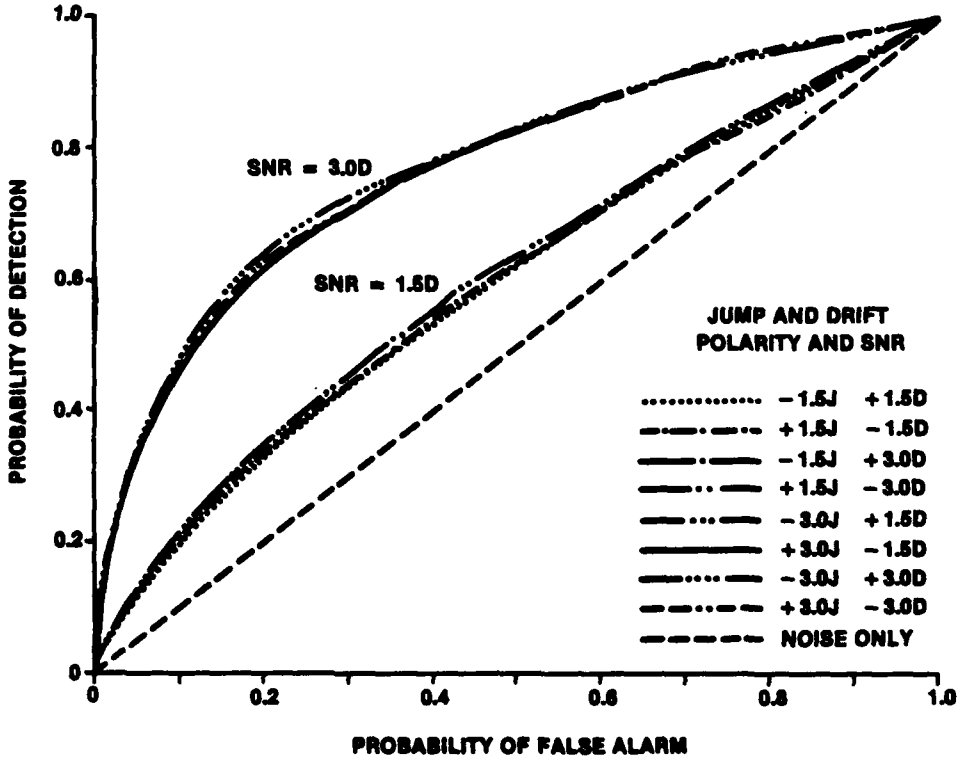


Figure 13d. χSQ Jump Output Given Drift with Opposite Polarity

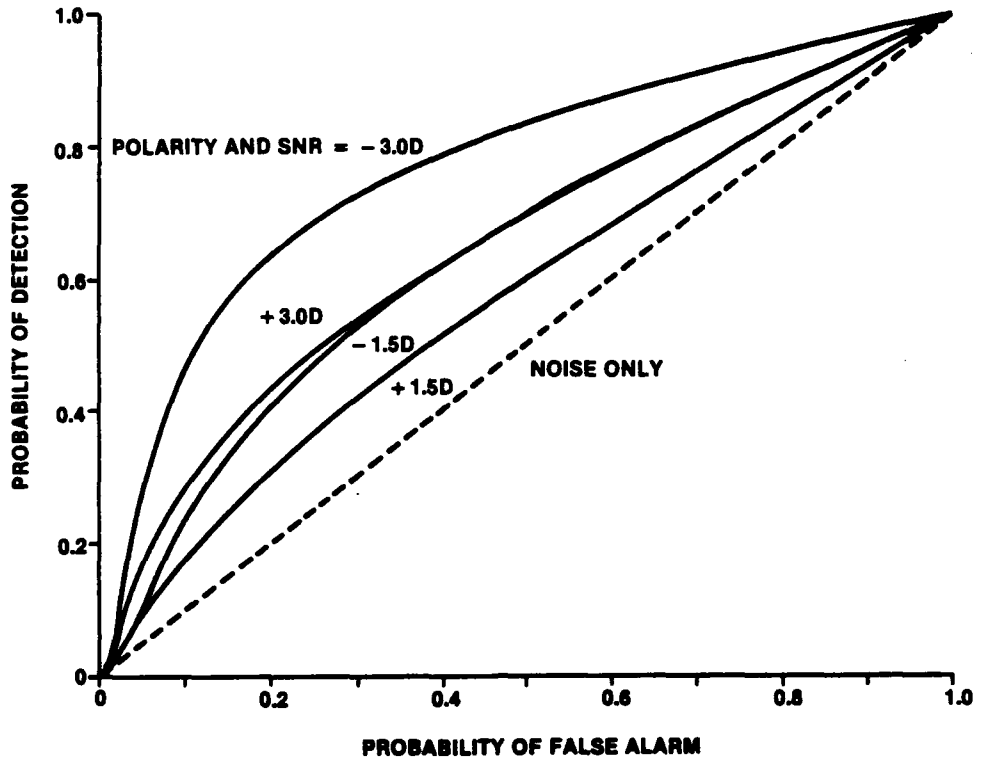


Figure 14a. ANN Jump Output Given Drift Only

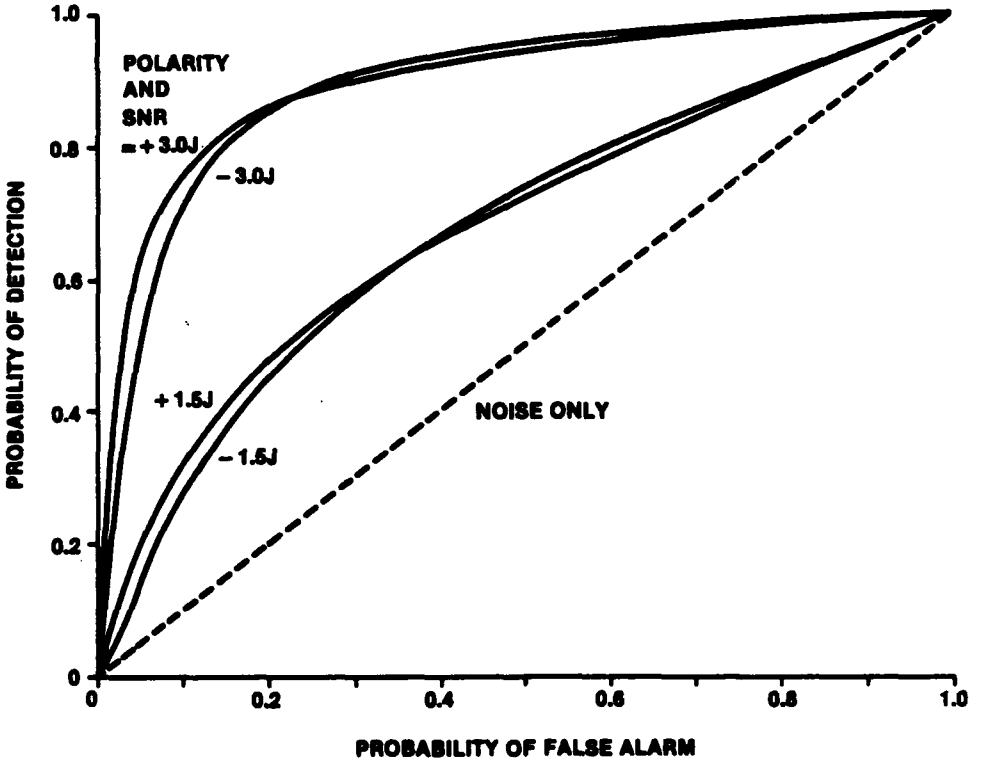


Figure 14b. ANN Jump Output Given Jump Only

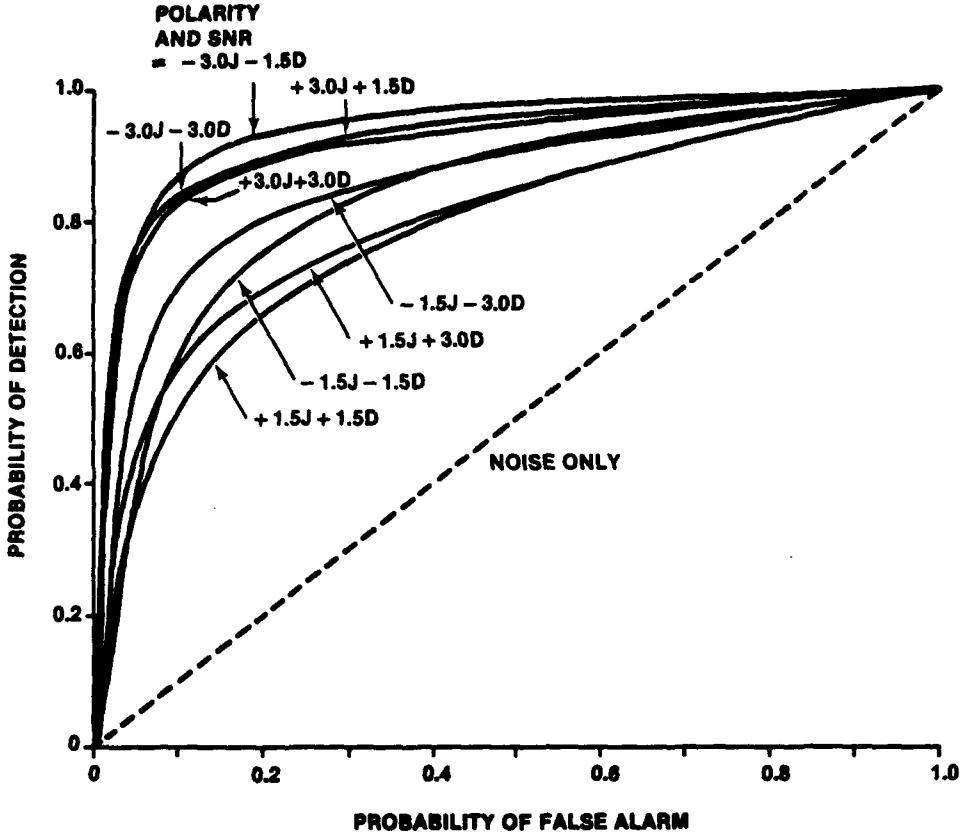


Figure 14c. ANN Jump Output Given Drift with Same Polarity

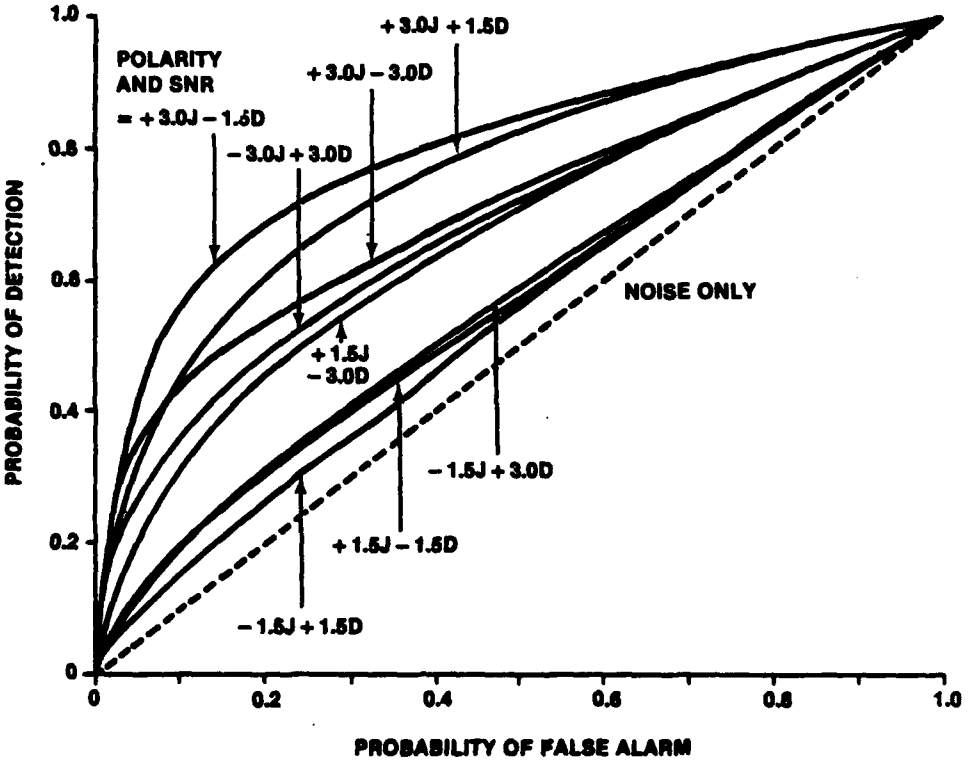


Figure 14d. ANN Jump Output Given Drift with Opposite Polarity

6. SUMMARY AND CONCLUSIONS

The problem of model assessment for nonlinear systems with uncertain models involves, among other things, detection and classification of mismodeling by observing residual sequences. Three methods for feature detection and extraction were developed and their performance investigated. The first method for multiple hypothesis testing was derived from a likelihood ratio test and resulted in a chi-squared (χSQ) test statistic. The second approach was arrived at through an *ad hoc* development that was designed to produce a similar test criterion. This approach resembles a modified Neyman-Pearson (MNP) test and is more computationally efficient than the χSQ test. The third method used an artificial neural network (ANN), which is a nontraditional approach to hypothesis testing. The ANN was trained to emulate the MNP test.

To facilitate analysis, the neural network was first trained and tested for the binary signal detection problem. Its performance was found to be equal to the theoretical bound of a generalized likelihood ratio test. For the multiple-feature application, the two classical techniques (a χSQ test and MNP test) were also presented for comparison. The signals considered were combinations of the different features of jump and drift at varying signal levels. Performance of the ANN relative to that of the other techniques was analyzed by comparing the percentage of detections of various feature combinations.

The results show that, although the ANN provides a somewhat better overall probability of detection (i.e., detection of jump with and without drift) at low probabilities of false alarm than the χSQ test, the apparent improvement in performance is tempered by the sensitivities of the ANN to an interfering signal. Viewing the constituents of the detection probability (i.e., jump detection without drift versus jump detection in the presence of drift) shows that feature detection is enhanced by the presence of an interfering signal with the same polarity, but it is inhibited by a signal with the opposite polarity. This is considered an undesirable effect. Hence, either an adequate network configuration or ample training of the network was lacking. However, because of the ANN's success with the binary hypothesis test, it is anticipated that further design modification can successfully eliminate the undesirable effects of the interfering signal, and that the ANN's performance can be made to match or exceed that of the χSQ and MNP techniques.

REFERENCES

1. J. Baylog, A.A. Magliaro, S.M. Zile, and K.F. Gong, "Underwater Tracking in the Presence of Modeling Uncertainty," *Proceedings of the Twenty First Asilomar Conference on Signals, Systems and Computers*, November 1987.

2. B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ, 1979.

3. H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, John Wiley and Sons, New York, 1968.

4. R. L. Plackett, *Regression Analysis*, Oxford University Press, New York, 1960.

5. C. R. Rao, *Linear Statistical Inference and Its Applications*, John Wiley and Sons, New York, 1965.

6. A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics, 5th Edition, Vol. 2.*, Oxford University Press, New York, 1991.

7. P. G. Hoel, S. C. Port, and C. J. Stone, *Introduction to Statistical Theory*, Houghton Mifflin, Boston, MA, 1971.

8. C. G. Y. Lau and B. Widrow, "Neural Networks I: Theory and Modeling (Scanning the Issues)," *Proceedings of the IEEE: Special Issue on Neural Networks*, vol. 78, no. 9, September 1990, pp. 1411-1413.

INITIAL DISTRIBUTION LIST

Addressee	No. of Copies
Chief of Naval Operations (OP-224)	1
Chief of Naval Research (OCNR-1271--J. Smith, OCNR-23--A. Faulstich, OCNR-232--D. Houser)	3
Naval Sea System Command (SEA-06UR, SEA-06UR3, SEA-06UR4)	3
Program Executive Officer, Submarine Combat and Weapons Systems (PMO-409--R. Dosti, PMO-418)	2
Commander, Submarine Force Atlantic Fleet	1
Commander, Submarine Force Pacific Fleet	1
Commander, Submarine Development Squadron Twelve	1
Naval Command and Control and Ocean Surveillance Center (Code 57--R. Moore)	1
Navai Postgraduate School	1
Naval Surface Warfare Center (Code U04--M. Stripling)	1
Defense Advanced Research Projects Agency	1
Defense Technical Information Center	2
Center for Naval Analyses	1